

**GENDER DIFFERENCES IN VERBAL, MATHEMATICAL, AND OVERALL
ACADEMIC SELF-CONCEPT: A META-ANALYTIC REVIEW**

A Thesis
Presented to
The Academic Faculty

by

Chelsea V. Velez

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
December, 2017

COPYRIGHT © 2017 BY CHELSEA V. VELEZ

**GENDER DIFFERENCES IN VERBAL, MATHEMATICAL, AND OVERALL
ACADEMIC SELF-CONCEPT: A META-ANALYTIC REVIEW**

Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Howard Weiss
School of Psychology
Georgia Institute of Technology

Dr. Rustin Meyer
School of Psychology
Pennsylvania State University

Date Approved: December 04, 2017

ACKNOWLEDGEMENTS

Thank you to my committee members, especially Dr. Richard Catrambone, for their continued guidance and support throughout this project. I would also like to thank Elnora Kelly and Victor Ellingsen for being the best friends and role-models I could have asked for in graduate school. Finally, thank you to my husband, Andrew Velez, for his never-ending love and patience.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iiii
LIST OF TABLES	vii
LIST OF FIGURES	vii
SUMMARY	viii
CHAPTER 1: Introduction and Literature Review	1
1.1 Differentiation of Self-Concept and Related Constructs	2
1.2 Development of Self-Concept Theoretical Models	5
1.3 Self-Assessed Intelligence and Lay Definitions of Intelligence	9
1.4 Gender Differences in Academic Self-Concept	11
1.5 Theories of Gender Differences and Similarities	13
1.6 The Current Study	17
1.7 Overall Relationship	18
1.8 Influence of Moderator Variables	19
1.8.1 Bem's Sex-Role Orientation	19
1.8.2 Type of Measure	20
1.8.3 Year of Publication	21
1.8.4 Sex of Authors	21
1.8.5 Nationality	22
1.8.6 Age	22
1.9 Objective Measures of Intelligence	23
CHAPTER 2: Method	24
2.1 Literature Search and Study Selection	24
2.2 Coding Studies	25
2.3 Inter-rater Reliability	28
2.4 Statistical Analyses	32
CHAPTER 3: Results	31
3.1 Differentiation of Self-Concept and Related Constructs	31
3.1.1 Magnitude of Gender Differences	31
3.1.2 Hypothesized Moderators	32
3.2 Verbal Self-Concept	33
3.2.1 Magnitude of Gender Differences	33
3.2.2 Hypothesized Moderators	34
3.3 Mathematical Self-Concept	34
3.3.1 Magnitude of Gender Differences	34
3.3.2 Hypothesized Moderators	35
3.4 Exploratory Analyses	35
3.4.1 Age, Author Gender, and Country of Origin	35
3.4.2 Objectively Measured Intelligence	36

CHAPTER 4: Discussion	37
4.1 Hypothesized Moderator Variables	37
4.2 Exploratory Moderator Variables	42
4.3 Objective Intelligence	43
4.4 Summary	45
4.5 Implications	46
4.6 Limitations	46
4.7 Conclusion	47
APPENDIX: Articles Used for Meta-Analyses	49
REFERENCES	58

LIST OF TABLES

Table 1. Summary of Meta-Analysis Results, as Moderated by Number of Items and Publication Year.....	31
Table 2. Summary of Meta-Analysis Results, as Moderated by Categorical Exploratory Sample Characteristics.....	44
Table 3. Summary of Meta-Analysis Results, as Moderated by Age	45

LIST OF FIGURES

Figure 1. Representation of self-belief constructs as a hierarchical model	4
Figure 2. Hierarchical representation of self-concept.....	7

SUMMARY

The literature provides many examples suggesting that males have, on average, higher academic self-concept levels when compared to females, especially in mathematical domains. Three meta-analyses were conducted to examine the magnitude of differences in overall academic, verbal, and mathematical self-concept. For all three constructs, males scored significantly higher than females on measures of self-concept, with weighted mean effect sizes from $d = 0.13$ to $d = 0.41$. Additionally, moderator variables such as the publication year of the study, were examined in an attempt to identify potential sources of gender differences in academic self-concept domains. The results of moderator analyses were varied across overall academic, verbal, and mathematical self-concept.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

With literature dating back to William James' 1890 publication, self-concept research has been an area of investigation for over a century. However, the 80 years following James' publication were filled with inconsistent forms of measurement and disparate definitions of self-concept (Byrne 1996). Shavelson, Hubner, and Stanton's (1976) critical evaluation of the self-concept literature marked a shift in the validation and exploration of the construct. The construct has gained much attention in the past 40 years, and the model proposed by Shavelson et al. has been greatly clarified since its publication (Marsh, 1990b; Marsh, Byrne, & Shavelson, 1988; Marsh & O'Neill, 1984). Seen as forces that guide human behavior, self-views are important when predicting general and specific outcomes including depression (Swann, Chang-Schneider, & McClarty, 2007), goal pursuit (Stipek & Gralinski, 1996), and academic performance (Hattie, 1992; Kornilova, Kornilov & Chumakova, 2009). Various factors such as gender and personality can moderate an individual's self-concept and related outcomes (Beloff, 1992; Furnham & Buchanan, 2005; Furnham, Shahidi, & Baluch, 2002).

My review will begin with a discussion of the operationalization of self-concept, including an examination of operational shortcomings and comparisons between related self-constructs. I will then summarize the development of the self-concept literature, with a primary focus on academic self-concept; this will include common measures used to assess academic self-concept. Finally, my review will address the gender differences in academic self-concept, and some of the theories commonly used to explain the gender

difference. Based on standards used in current gender difference research, the following guidelines will be used to evaluate effect sizes: $d = 0.20-0.49$ considered a small difference, $d = 0.50-0.79$ considered moderate, and d greater than 0.80 considered large (Cohen, 1988; Hyde, 2005). Additionally, $d < 0.20$ will be considered a trivial sex difference.

1.1 Differentiation of Self-Concept and Related Constructs

One of the primary concerns of self-concept research is the lack of a consistently used operational definition (Byrne, 1984). Although there have been numerous attempts to explicitly define the construct of self-concept, the findings have been inconclusive (e.g., Byrne, 1984; Hansford & Hattie, 1982; Shavelson et al., 1976). For example, Shavelson et al. (1976) determined 17 different dimensions on which the construct could be organized. Also, due to the broad definition of self-concept, researchers have tended to use certain terms synonymously and without conceptual clarification (Hattie, 1992). It is important to note that, within the same article, self-concept may also be labeled interchangeably as *self*, *self-estimation*, and/or *self-perception* (Byrne, 1996).

Along with using terms synonymously throughout the literature, the self-concept literature suffers from imprecise theoretical definitions and inconsistent operationalization. Marsh (1990a) explained that “self-concept, like many other psychological constructs, suffers in that ‘everybody knows what it is,’ so that many researchers do not feel compelled to provide any theoretical definition of what they are measuring” (p. 79). Therefore, it is important to explicitly distinguish self-concept from other self-beliefs to ensure that the measurement and theoretical model being used are cohesive. Self-esteem and self-efficacy are related constructs that are commonly confused with or poorly differentiated from self-concept. This is, in part, due to the constructs correlating so highly with each other that

researchers have questioned whether they are truly different (for a review, see Pajares, 1996). For example, when examining domain specific self-beliefs, Pajares and Miller (1994) found that mathematical self-efficacy and mathematical self-concept were highly correlated ($r = .61$). Additionally, Chen, Gully, and Eden (2004) found that measures of general self-efficacy and measures of self-esteem were also highly correlated ($r = .59-.74$). Though the constructs are similar and related, it is essential to recognize their theoretical distinctions (Valentine, DuBois, & Cooper, 2004).

When looking at self-beliefs, it is helpful to consider the related constructs as a hierarchical model (see Figure 1). Self-esteem, sometimes referred to as “global self-concept,” is at the apex of the structure (Marsh & Yeung, 1998; Trautwein, Lüdtke, Köller, & Baumert, 2006), and refers to an individual’s overall feeling of self-worth (Bong & Clark, 1999; Hattie, 1992; Rosenberg, 1965). Self-esteem consists of appraisals of the descriptive components of self-concept (Valentine et al., 2004) and is high in affective evaluation (Byrne, 1996; Marčič & Kobal Grum, 2011). Focus on self-worth can be seen in items found on the Rosenberg Self-Esteem Scale such as “I take a positive attitude toward myself” and “In general, I am satisfied with myself” (Rosenberg, 1965).

In contrast, self-efficacy is measured with the greatest level of specificity (Pajares & Miller, 1994), and can be described as “people’s judgements of their capabilities to organize and execute courses of action required to obtain designated types of performances” (Bandura, 1986, p. 391). Self-efficacy is context-dependent and focuses on what one expects he/she can achieve in terms of specific tasks in given situations (Bong & Skaalvik, 2003; Pajares & Miller, 1994). Measures of self-efficacy are concerned with how individuals view their potential to accomplish tasks and ask questions relating to a specific

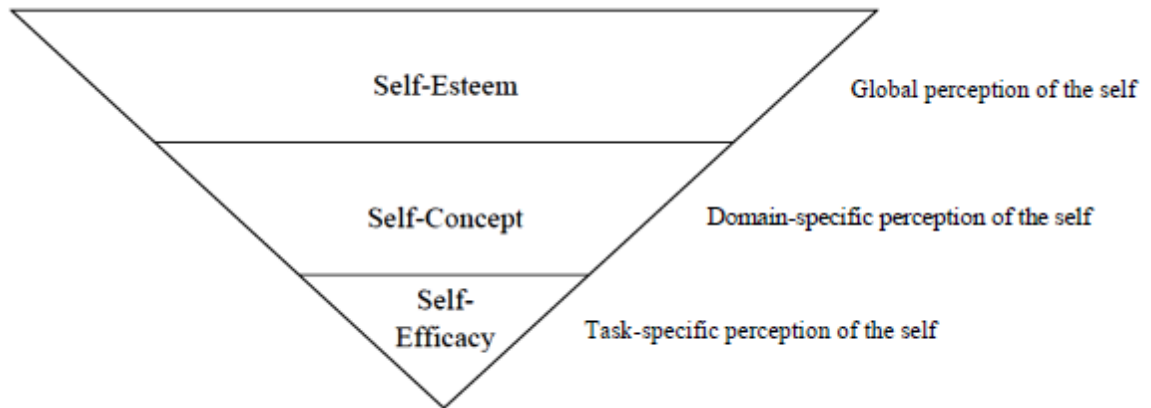


Figure 1. Representation of self-belief constructs as a hierarchical model. The model consists of self-esteem positioned at the top, domain-specific self-concept in the middle, and situational evaluations of self-efficacy at the base. Created based on “Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept.” by Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). *Journal of Personality and Social Psychology*, 90, 334-349.

action under certain circumstances (Bong & Skaalvik, 2003). For example, self-efficacy measures might include items such as “How sure are you that you can *write a simple sentence with good grammar?*” (Pajares, Miller, & Johnson, 1999).

Self-concept lies somewhere in between the previous constructs in the self-belief hierarchy; it is not as global as self-esteem, yet it is not task-specific like self-efficacy. Defined most broadly as “a person’s perception of himself,” self-concept tends to be domain specific (Shavelson et al., 1976, p. 411). Additionally, compared to self-esteem, which is primarily affective, and self-efficacy, which is primarily cognitive, self-concept is composed of cognition, affect, and behavior (Byrne, 1996; Bong & Skaalvik, 2003). Self-concept judgements rely on how competent a person believes he/she is in a certain domain (e.g. “I am a good science student”), rather than assessing overall self-worth or ability to perform task-specific behaviors (Bong, Skaalvik, 2003; Pajares & Miller, 1994).

1.2 Development of Self-Concept Theoretical Models

With little uniformity in the use of an operational definition and the challenging differentiation of self-belief constructs, the self-concept literature has a history of being indefinite. For the 80 years following James’ (1890) work, the conceptualization of self-concept was inconsistent, and researchers struggled with comparing results and interpreting findings due to varying operational definitions (Shavelson et al. 1976). To rectify these issues, Shavelson et al. (1976) attempted to operationally define self-concept in hopes of creating a cohesive, overarching theory for researchers to work from. Through their investigation of five major instruments, they identified seven components that were crucial to defining the construct; these components were that self-concept is: (a) organized,

(b) multifaceted, (c) hierarchical, (d) stable, (e) developmental, (f) evaluative, and (g) differentiable (Shavelson, et al., 1976).

Shavelson et al. (1976) suggested that one's self-concept is formed through experience with the environment and feedback from significant others; also, self-concept is both cause and effect in that it influences behaviors, which in turn can influence one's self-concept. These authors also proposed that self-concept is a multidimensional and hierarchically ordered structure with general self-concept positioned at the zenith and evaluations of behaviors in specific situations at the bottom (see Figure 2). In the middle of the proposed hierarchical model, self-concept is then divided into two primary facets: academic and non-academic. The second-order facets of self-concept are also subdivided into four domains within academic self-concept: English, history, math, and science (Shavelson, et al., 1976). The non-academic component of the model fell outside the area under investigation. Although Shavelson et al. did not initially justify the model with empirical evidence, their posited model is one of the most comprehensively validated models of self-beliefs and was a milestone in the renaissance of self-concept research (Byrne, 1996; Marsh, 2006).

In hopes of addressing the lack of a sufficient measure of self-concept, Marsh and colleagues developed the Self-Description Questionnaire (SDQ) instruments (Marsh & O'Neill, 1984). Based primarily on the theoretical foundation of Shavelson et al. (1976), Marsh and O'Neill (1984) designed the SDQ-III measure to examine self-concept in late adolescents using the following domains: mathematics, verbal, academic, problem-solving, physical abilities, physical appearance, same-sex peers, opposite-sex peers, relations with parents, religion/spirituality, honesty/reliability, emotional stability, and general self

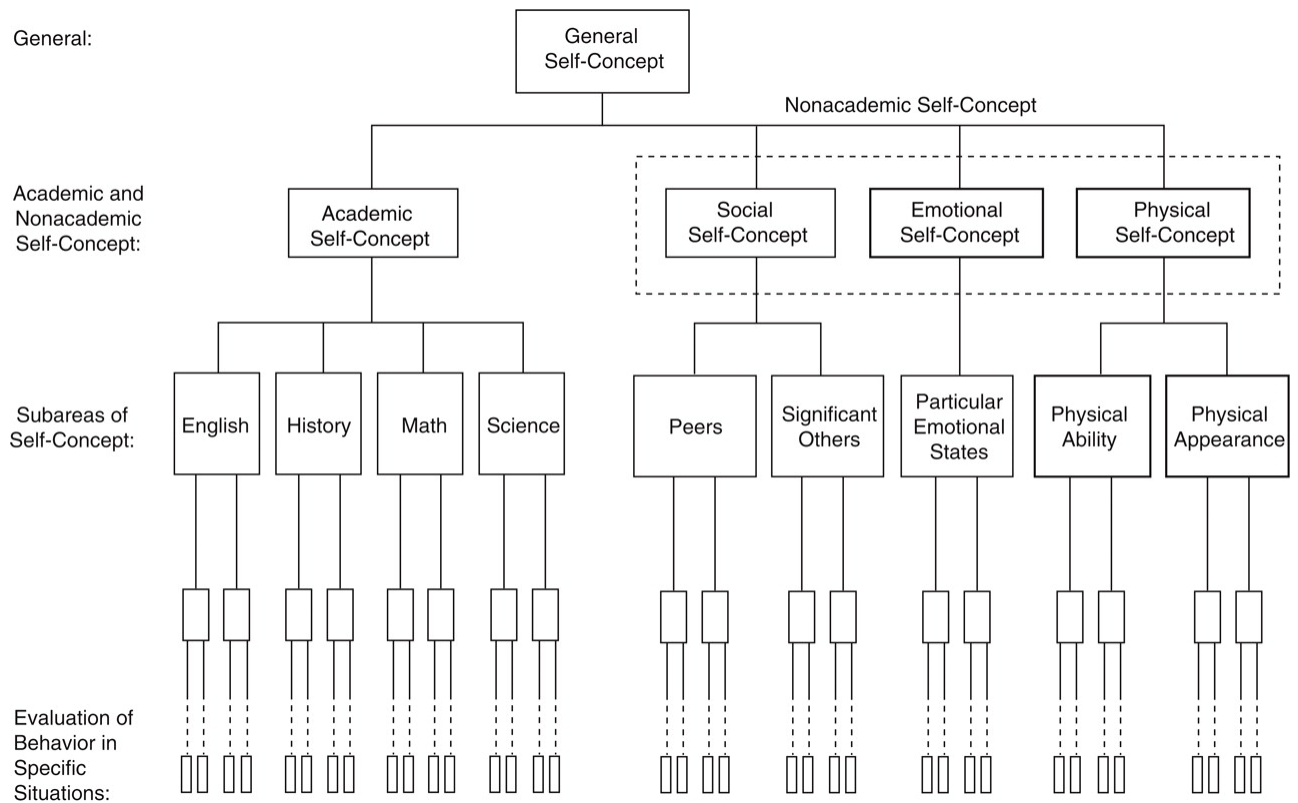


Figure 2. Hierarchical representation of self-concept with general self-concept positioned at the apex and situation-specific evaluations of behaviors at the base. Adapted from “Self-Concept : Validation of Construct Interpretations,” by Shavelson, R. J.

(Marsh & O'Neill, 1984). Each of the subscales measuring the different domains in the SDQ-III consists of 10 or 12 items with 8-point Likert-type response scales (Marsh, 1992). While investigation of the SDQ-III provided support for the multidimensionality of self-concept (domain correlations had a mean $r = .08$), the hierarchical structure and the correlations between domains were weaker than expected with nothing significantly correlating with the general self factor (e.g. Marsh, 1990b; Marsh, 2006; Marsh & O'Neill, 1984; Marsh, Byrne, & Shavelson, 1988).

Though some research supported the Shavelson et al. (1976) model (e.g., Shavelson & Bolus, 1982), subsequent research showed that math and verbal self-concept factors do not correlate highly enough to be considered a single higher-order academic self-concept (Marsh, 1987; Marsh, 2006; Marsh & O'Neill, 1984). Despite the overall hierarchy of the Shavelson model being weak, Marsh & O'Neill recognized that academic scales could in fact be represented on two higher order factors: verbal ability and math ability. Interestingly, during the process of validation and reliability testing, although math and verbal achievement were highly correlated with each other ($r = .59$) and self-concept in each domain was highly correlated with achievement in the same domain (math $r = .58$, verbal $r = .42$), math and verbal self-concept were uncorrelated with each other ($r = -.04$; Marsh & O'Neill, 1984).

In response to the surprisingly low correlations between math and verbal self-concepts, Marsh and colleagues proposed two different theoretical advancements. The first addition Marsh made to the literature was devising a model related to Shavelson's, but with more specificity in academic self-concept (Marsh & Shavelson, 1985; Marsh, Byrne, & Shavelson, 1988). After testing competing models, Marsh and colleagues identified that

having two higher-order factors (math and verbal self-concept) fit the data significantly better than the original model with only a single, overall higher-order factor (Marsh & Hocevar, 1985; Marsh & Shavelson, 1985). This new model, known as the Marsh/Shavelson revision, was supported by subsequent research by Marsh and his colleagues (Marsh, 1987; Marsh, et al., 1988; Marsh, 1990b).

Second, in regards to the uncorrelated math and verbal self-concept measures, Marsh (1986) suggested that people use two different frames of reference when forming self-concepts; this theory became known as the internal/external frame of reference model. Specifically, people use an external frame of reference where they compare their abilities to the abilities of others, and an internal frame of reference where they compare their level of ability in one domain to their level of ability in another. Because objective math and verbal abilities are substantially correlated with each other ($r = .5$ to $.8$), external comparisons lead to a positive relationship between math self-concept and verbal self-concept (Marsh, 1986). Conversely, internal comparisons amplify the intraindividual differences between math and verbal abilities, resulting in a negative relationship between math self-concept and verbal self-concept. When taken together, these comparisons end up negating each other, resulting in a near-zero correlation between math and verbal self-concept ($r = -.03$ to $.03$; Marsh, 1986).

1.3 Self-Assessed Intelligence and Lay Definitions of Intelligence

In addition to self-concept measures, researchers also have used self-assessed intelligence (SAI) measures to evaluate self-perceptions of cognitive ability. Although some researchers have advocated for the differentiation of self-estimation and academic self-concept as constructs (Freund & Kasten, 2012), there is considerable overlap in their

conceptualizations (Peterson & Whiteman, 2007). Self-estimates are understood as expressions of self-concept, and are vital components of a person's overall self-view (Epstein, 1973; Freund & Kasten, 2012). Additionally, SAI tends to focus on people's lay perceptions of intelligence (Chamorro-Premuzic & Furnham 2005; Furnham, 2001). Compared to *explicit* theories of intelligence, which refer to those that are based on collected cognitive function data, SAI is grounded in *implicit* theories, which refer to ideas that are constructed by individuals (i.e. lay persons) based on their experiences (Freund & Kasten, 2012; Sternberg, 1990). Though there has been little empirical evidence supporting the theory (Furnham, Hosoe, & Tang, 2002), Gardner's (1983, 1999) notion of multiple intelligences is commonly used to assess SAI due to the assumption that lay people's implicit theories of intelligence are generally multidimensional (Chamorro-Premuzic & Furnham, 2005).

Gardner's (1983) pluralistic approach to intelligence stemmed from his belief that traditional measures of academic achievement were too restrictive and relied too heavily on typical academic domains of linguistics and mathematics. He posited that all individuals possess a level of intelligence in eight different domains, and one's intelligence "shape" is made up of varying strengths and weaknesses across the different areas (Gardner, 1999). Therefore, when obtaining participants' SAI, it is common for surveys to request rankings for each domain, as well as a rating for overall intelligence. Though Gardner (1983) did not explicitly operationally define any of the intelligences, based on his provided parameters, researchers have defined his verbal domain, *linguistic intelligence*, as "sensitivity to spoken and written language and the ability to learn languages," and his math domain, *logical/mathematical intelligence*, as "the capacity to analyze problems

logically, solve math problems and investigate issues scientifically” (Furnham & Fukumoto, 2008, p. 64). Self-estimates are typically measured by having participants rate their intelligence based on normal distribution curves with labeled standard deviations and means (Furnham & Baguma, 1999; Peterson & Whiteman, 2007). However, it should be noted that when self-concept measures are administered in a format similar to self-estimations, they have generally the same pattern of validity correlations with other self-estimation measures (Ackerman & Wolman, 2007).

1.4 Gender Differences in Academic Self-Concept

The academic self-concept literature includes examination of many other variables, but the most prominent and frequently discussed topic is gender (Hattie, 1992). Though there is some disagreement among researchers (Skaalvik, 1997; Stake, 1992), men’s self-estimates of overall intelligence are commonly higher than women’s (Bennett, 1996; von Stumm, Chamorro-Premuzic, & Furnham, 2009). Some researchers have hypothesized that the difference in self-report measures between genders is due to men’s overestimation and women’s underestimation of their own intelligence. This is often referred to as ‘male-hubris – female humility,’ and has been studied across various cultures (Furnham, Hosoe, & Li-Ping Tang, 2002; Kaufman, 2012; Syzmanowicz & Furnham, 2011). Hogan (1978) was one of the first researchers to examine the differences between men’s and women’s perceptions of their own and others’ levels of intelligence, and found that, compared to men, women consistently underestimated their own intelligence. These findings were supported by Beloff (1992), and subsequent research on self-concept has primarily found significant gender differences in self-reported overall intelligence (e.g. Furnham & Rawles, 1995; Marsh, 1994; Workman, 2004).

In addition to overall intelligence, gender differences have been examined in more specific domains of self-concept such as verbal and mathematical ability. In general, mathematics is stereotyped as a male domain, whereas verbal ability is stereotyped as a female domain (Guimond, Chatard, Martinot, Crisp, & Redersdorff, 2006; Nosek et al., 2009). For example, when estimating their children's scores, parents anticipated higher scores for boys compared to girls on mathematical ability, and higher scores for girls compared to boys on verbal ability (Furnham & Bunclark, 2006). When gender differences are found in verbal and mathematical self-concept, they tend to correspond with traditional gender-roles and stereotypes (Skaalvik & Skaalvik, 2004). A recent meta-analysis of SAI found the largest weighted mean gender differences effect size was for mathematical–logical self-concept ($d = .44$), suggesting that men had significantly higher self-estimations in the domain than women (Syzmanowicz & Furnham, 2011). Additionally, a meta-analysis of gender differences among children and adolescents suggests a small, but significant, gender difference in mathematical self-concept ($d = .29$; Wilgenbusch & Merrell, 1999). Though there have been some conflicting results in the literature (for contradictory findings see Furnham and Akande, 2004), recent meta-analyses support the notion that men rate themselves significantly higher than women do on mathematical abilities.

Also, in line with traditional gender roles, self-perception of ability in verbal domains appears to be higher for females than for males, when there is a difference (Furnham, Rakow, Sarmany-Schuller & DeFruyt, 1999; Nasser & Singhal, 2006). Wilgenbusch and Merrell (1999) found a small, yet significant effect size favoring females

in their meta-analysis ($d = -.23$). However, Syzmanowicz and Furnham (2011) found a contrasting result of a marginal gender difference favoring men ($d = .07$).

1.5 Theories of Gender Differences and Similarities

In general, there are two competing views explaining gender differences in self-concept: theories that suggest an underlying gender difference in actual, psychometrically tested ability due to biological differences, and theories that suggest differences due to gender stereotypes and magnitude of sex-role orientation (Freund & Kasten, 2011). I will discuss the two major competing theoretical categories below.

Though not the first to study gender differences in academic self-concept, Beloff (1992) was the first to address the trend for women to rate their own intelligence significantly lower than men rated theirs. Replicating Hogan's (1973) findings that women tended to underestimate their IQ scores, Beloff (1992) argued that these self-estimates of ability were not representative of the sample's actual cognitive ability and were partially due to gender stereotypes. Campion (1992) countered that there was no reason to assume that the two groups were of equal intelligence, and therefore, the assumptions made by Beloff were unwarranted due to objective intelligence not being measured. Nevertheless, in regards to self-estimates of overall intelligence, an underlying gender difference in actual cognitive performance due to biological differences does not seem to be a comprehensive explanation (Beyer, 1990). Beyer found that when evaluating gender differences in performance expectancies, actual performance, and self-evaluations of stereotypically gender-typed tasks, expectancies affected self-evaluations above and beyond actual performance. This suggests that objective cognitive ability is not the only factor underlying the development of an individual's self-concept (Beyer, 1990).

The current consensus in the literature is that there are no differences between men and women on measures of overall intelligence (Colom, Juan-Espinosa, Abad, & Garcia, 2000). However, it is important to identify how overall intelligence is being conceptualized when considering gender differences in ability. With differences varying in magnitude for various domains, the difference between men and women on a general intelligence measure can vary depending of the content of the test (Ackerman, 2006). For example, when examining the differences between estimated intelligence and objective intelligence for men and women, Reilly and Mulhern (1995) only used the Digit Symbol and Vocabulary tests from the WAIS to extrapolate participants' IQ scores. Though these subscales might have been satisfactory measures of objective overall intelligence, the results could have been very different if they used a different measure such as the Stanford-Binet IQ test. Without an overall consensus of how to define general intelligence, researchers must be explicit in their construct definitions when examining gender differences in overall intelligence (Ackerman, 2006).

Though specific domains of intelligence might seem easier to operationalize than general academic ability, mathematical intelligence and verbal intelligence findings are actually more complicated. For mathematical ability, meta-analyses suggest that gender differences in favor of men are trivial in magnitude (Lindberg, Hyde, Petersen, & Linn, 2010), but can increase in magnitude when looking at complex problem solving (Hyde, Fennema, & Lamon, 1990). Conversely, for verbal ability, studies have suggested minimal gender differences favoring women (Hyde, 2014; Hyde & Linn, 1988; Wilgenbusch & Merrell, 1999). For both objective math and verbal ability, the magnitude of the gender differences continues to be a topic of debate. It is possible that part of the discrepancy in

results is due to varying levels of specificity or cognitive level in assessments used across studies (Hyde, Fennema, & Lamon, 1990). For overall intelligence and specific domains, it is still unclear whether or not gender differences in self-concept reflect actual differences in objectively measured intelligences. Although the best way to investigate the extent to which gender differences on subjective measures reflect differences on objective measures would be a direct comparison with the same samples, there are surprisingly few studies addressing the issue (Freund & Kasten, 2011).

In spite of objective measures of intelligence generally supporting Hyde's (2005) hypothesis that males and females are more alike than they are different, gender differences persist in self-concept literature. Where do these gender differences come from? According to cognitive social learning theory, both children and adults attempt to emulate others in their environment, and learn via reinforcements and punishments (Bussey & Bandura, 1999). At a very young age, children internalize cultural stereotypes and conform to societal expectations of gender, and this adherence to gender norms is further reinforced by parents' differential treatment of their sons and daughters (Eagly & Wood, 2013). Bong and Skaalvik (2003) explained that people view themselves the way others perceive them; this follows Shavelson et al.'s (1976) idea that self-concept is formed through experiences with the environment and significant others. Furthermore, Gentile, Grabe, Dolan-Pascoe, Twenge, Wells, and Maitino (2009) suggested that self-evaluations are influenced by a "generalized other," which leads to differential expectations for men and women.

Another explanation for gender differences in self-concept is Eagly and Wood's (1999; Wood & Eagly, 2012) sociocultural theory. This theory suggests that society's division of labor by gender underlies all other psychological discrepancies (Hyde, 2014).

Therefore, gender differences stem from a person's adaptations to the various limitations on or opportunities for their gender. In support of sociocultural theory, Nosek et al. (2009) found that nations' sex differences in science and mathematics achievement were significantly correlated with the nations' implicit gender-science stereotyping. Whether through sociocultural or cognitive social learning theory, gender stereotypes and gender role expectations exist. Males are more commonly ascribed intellectual competence than females (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972), and Furnham (2000, 2001) has argued that intelligence is male normative. Specifically, mathematical intelligence is viewed as more masculine, and verbal intelligence is seen as more feminine (Bennett, 2000).

Though boys and girls differ in early socialization, there is variation in the degree to which individuals develop stereotypically gendered traits (Reilly & Neumann, 2013). Bem (1981) posited that self-concept is directly related to people's gender schema, and through learning and schematic processing, self-concept becomes sex-typed over time. She suggested that, due to internalized sex-type, people regulate their behavior so that it conforms to society's cultural definitions of 'maleness' or 'femaleness.' This adherence to traditional sex-roles shapes both self-concept and behavior, and through this process, gender stereotypes become a self-fulfilling prophecy. Bem also suggested that the more sex-typed a person is (i.e. how strongly they identify themselves as their gender), the more likely they are to adhere to traditional gender roles (Bem, 1981).

Gender stereotypical traits and behaviors influence academic self-concept in many ways. Roberts (1991) suggested that males and females interpret information related to failure/success differently. Women tend to take note of both positive and negative

commentary, whereas men overlook negative information and retain only positive feedback. This difference in reception and recall of feedback potentially inflates male self-evaluation (Beloff, 1992). Syzmanowicz and Furnham (2013) suggested that, rather than gender differences arising due to females tending to be more self-deprecating, the discrepancy in self-concepts is due to a self-enhancing effect of masculinity.

1.6 The Current Study

Several factors warrant a new meta-analytic investigation of gender differences in academic self-concept. The most recent meta-analysis that focused on self-perceptions of academic ability was Syzmanowicz and Furnham's (2011) study on self-estimates. Though this analysis is relatively recent and provided an overview of current research, the article was not without limitations. Syzmanowicz and Furnham's studies yielded what they considered moderate effects for general intelligence ($d = 0.37$), mathematical intelligence ($d = 0.44$), and spatial intelligence ($d = 0.43$); minimal effects were found for verbal intelligence ($d = 0.07$). All results favored men. However, due to their position that *academic self-concept* and *self-assessed intelligence* were different constructs, Syzmanowicz and Furnham did not utilize any studies of academic self-concept. In addition to the analysis being limited to only self-estimates of ability, the article used the fixed-effect method which tends to inflate Type I error rates, and relied primarily on an abundance of single-variable one way ANOVAs to assess moderators without controlling for their multiple comparisons. The Syzmanowicz and Furnham (2011) article provided a framework from which to build, but the scope and methods leave room for improvement.

I will conduct three meta-analyses, examining gender differences in verbal, mathematical, and overall academic self-concept. The study will use a collection of articles

on both academic self-concept and self-estimates of intelligence to provide a more accurate assessment of the size and significance of gender differences in self-perceptions. In addition to considering studies using self-concept measures and self-estimate measures, the study will investigate key moderators in hopes of further clarifying the current literature.

1.7 Overall Relationship

As previously discussed, most studies investigating overall academic self-concept have found significant differences favoring men. Although the finding of gender differences has been consistent, the magnitude of the effect has varied. Based on Syzmanowicz and Furnham's (2011) recent findings, I expect to find a moderate difference ($d = 0.5-0.79$) favoring men on overall academic self-concept. Following trends in previous meta-analyses, I anticipate that men rate themselves significantly higher than women on mathematical self-concept (Syzmanowicz and Furnham, 2011; Wilgenbusch & Merrell, 1999). Finally, I expect trivial gender differences in verbal self-concept.

A power analysis equation for main effects using mixed-effects models could not be found. Data from Syzmanowicz and Furnham's (2011) meta-analysis were used to approximate power using fixed-effect method power analysis. Using sample sizes and effect sizes from the 2011 article, power for an effect size of .10 was greater than 80% for mathematical, verbal, and overall academic self-concept at the $\alpha = .05$ level.

Hypothesis 1: The first meta-analysis will reveal a significant gender difference in overall academic self-concept, favoring males. It is expected that this effect size will be moderate ($d = 0.50-0.79$).

Hypothesis 2: The second meta-analysis will reveal a significant gender difference in mathematical self-concept, favoring males. It is expected that this effect size will be moderate ($d = 0.50-0.79$).

Hypothesis 3: The third meta-analysis will reveal trivial gender differences in verbal self-concept, favoring males ($d < 0.10$).

1.8 Influence of Moderator Variables

In addition to examining the main effects for gender differences in mathematical, verbal, and overall academic self-concept, moderator variables will be considered. Based on sociocultural theory and cognitive social learning theory, I expect sex-role orientation to significantly moderate the magnitude of gender differences in self-concept. Also, due to sex-role expectations and societal norms changing over time, I expect date of publication to be a significant moderator as well. Finally, due to men's tendency to make self-enhancing evaluations when given ambiguous assessments, I expect explicit measures of self-concept domains such as the SDQ-III to have lesser gender differences. I discuss each of these moderators in greater detail below.

1.8.1 Bem's sex-role orientation

Many researchers have used Bem's Sex Role Inventory (BSRI), which is grounded in Bem's gender schema theory, to measure the potential moderating effect of gender role on academic self-concept (Erdwins, Small, & Gross, 1980; Storek & Furnham, 2012; Syzmanowicz & Furnham, 2013). The BSRI is a measure of psychological sex-role orientation. Subjects can be typed as feminine, masculine, androgynous (having both feminine and masculine qualities), or undifferentiated (possessing neither feminine nor

masculine traits; Bem, 1974). Bem (1981) posited that the more sex-typed a person is (i.e. being feminine or masculine), the more likely he/she is to adhere to stereotypical gender roles. Rammstedt and Rammsayer (2002) supported this theory, and found that men's overestimation of their intelligence is relative to the strength of their adherence to gender roles. Few studies have examined the relationship between sex-type and gender differences in self-concept, and findings have been inconsistent (Syzmanowicz & Furnham, 2013). However, based on Bem's gender schema theory and previous research on gender roles, I expect that gender-typed people (masculine males and feminine females) will have greater gender differences than non-gender-typed people (both/neither sex typed: androgynous/undifferentiated).

Hypothesis 4: The magnitude of gender differences will be moderated such that samples of gender-typed men and women will have larger gender differences in all three self-concept domains than samples of non-gender-typed men and women.

1.8.2 Type of measure

In general, people assess themselves more positively than they assess other people (Brown, 2012). This tendency for positive evaluation on different traits and abilities (known as the 'better-than-average effect'; Alicke, 1985) is strengthened when the trait being evaluated is ambiguous (Dunning, Meyerowitz, & Holzberg, 1989; Felson, 1981). Men are particularly likely to make self-enhancing evaluations due to their tendency to recall only positive feedback (Thomas, 1991). When assessments are ambiguous, men are able to use multiple and variable evaluation criteria when estimating their intelligences, and are likely to recall positive experiences on which to base their estimations. Compared to women, who generally recall both positive and negative feedback, men's evaluations

become inflated (von Stumm, 2014). This ability to choose multiple criteria on which to base an estimation is exaggerated when the measure is ambiguous. Therefore, for verbal and mathematical self-concept, it is expected that gender differences will be exacerbated on ambiguous measures (identified by a single item questionnaire for each domain).

Hypothesis 5: Gender differences on verbal, mathematical, and overall academic self-concept will be larger for measures utilizing ambiguous questionnaires, compared to more thorough assessments such as the SDQ-III.

1.8.3 Year of publication

The year of publication should influence gender differences in academic self-concept because gender stereotypes have changed over time. Hattie (1992) suggested that self-concept changes over generations, and Shipstone and Burt (1973) suggested that stereotypes about intelligence may change over time as well. Due to men and women's roles becoming more comparable, it has been suggested that the magnitude of gender differences in various domains is decreasing (Byrne, Miller, & Schafer, 1999; Hyde, 2005). Therefore, studies from more recent decades are hypothesized to have smaller effect sizes.

Hypothesis 6: Year of publication will significantly moderate gender differences in all three meta-analyses so that differences decrease as studies become more recent.

The following variables will be investigated as moderators in an exploratory fashion.

1.8.4 Sex of authors

Recent meta-analyses found that male and female researchers tend to produce findings that shed a more favorable light on their own gender (Eagly & Wood, 2013). Gender composition of the authors was significantly related to the effect sizes in studies, with all-male research teams producing larger mean effect sizes than groups with women (Syzmanowicz & Furnham, 2011). Therefore, I expect larger sex differences in studies conducted by all or majority male authors, compared to research teams consisting of all or majority females.

1.8.5 Nationality

With research now being conducted cross-culturally, it is important to investigate variation across nations. Syzmanowicz and Furnham (2011) found a significant moderating effect of nationality, and after dividing studies into sub-samples, discovered that “Western” samples had significantly larger gender differences in mathematical self-estimates compared to non-Western samples. A significant interaction effect of sex by nation was found by von Stumm, Chamorro-Premuzic, & Furnham (2009) in their study across 12 nations. The authors suggested that individualistic societies might be more likely to overestimate their abilities.

1.8.6 Age

Though little research has examined how gender differences in self-concept vary with age, gender stereotype development research suggests that the acknowledgement and adherence to traditional gender roles increases with age (Kessels, 2005). Even when limiting the sample to elementary and middle school-aged children, 8th grade girls hold significantly stronger gender-stereotypical beliefs about math compared to 4th grade girls ($d = 0.37$; Kurtz-Costes, Copping, Rowley, & Kinlaw, 2014). The hypothesis that older

samples have larger gender differences than younger samples was supported by Syzmanowicz & Furnham (2011), and is expected in the current analysis as well.

1.9 Objective Measures of Intelligence

Furnham (2001) posited that gender differences in self-estimations of ability exist because of the “male hubris, female humility” effect. This effect suggests that men are likely to overestimate their intellectual abilities, whereas women are likely to underestimate their abilities. However, without a comparison of objective measures of intelligence, there is no way to know whether men and women are making inaccurate assessments. Ackerman and Wolman (2007) found that people are not accurate when assessing their abilities, and correlations between estimated and psychometrically tested intelligence scores rarely exceed $r = .50$. Studies have found gender differences in academic self-concept that exceed differences on psychometric intelligence (Furnham, Moutafi, & Chamorro-Premuzic, 2005), but the number of studies reporting this information is limited. When possible, I will obtain gender differences for studies examining objective measures and self-estimates of intelligence, and will compare the effect sizes of those gender differences.

CHAPTER 2

Method

2.1 Literature Search and Study Selection

To identify relevant studies, three primary strategies were used. First, computerized literature searches were conducted using primary psychology (PsychINFO and CogNet) and education (ERIC) databases, as well as more general research sources such as Google Scholar and Web of Science. To search the databases, the keywords *self-appraisal*, *self-assess*, *self-concept*, *Self-Description Questionnaire*, *self-estimate*, *self-report*, *self-evaluation*, *perceived ability*, and *perceived competence* were crossed with *academic intelligence*, *achievement*, *aptitude*, *intelligence*, and *multiple-intelligence* using the Boolean operator AND. Asterisks were used at the end of search terms to allow for a truncation. The database searches resulted in a total of 5,652 independent peer-reviewed articles for inclusion consideration. Attempts to reduce the number of irrelevant articles were made by removing generic search terms. However, due to the interchangeability of self-concept terms, it was critical to use the full array of search terms to capture pertinent articles.

The second strategy used to locate articles was to scan reference lists of the studies identified through the database searches. Finally, researchers currently conducting studies relating to academic self-concept were contacted to obtain any in-press or unpublished studies, as well as to request more specific data from published articles related to my hypotheses. No additional studies or data were received.

Articles from the database searches and reference scans were evaluated for inclusion in two phases. The initial phase required that the title or abstract alluded to measures of overall academic, verbal, or mathematical self-concepts. After removing articles with irrelevant titles or abstracts, 2,147 publications remained. These articles were read and evaluated based on the following criteria: (a) the study included both men and women; (b) the sample had a mean age of at least 16 years; (c) the study contained original data; (d) the study measured overall academic, verbal, or mathematical self-estimates or self-concepts; and (e) the article included statistics that enabled the calculation of appropriate gender comparison effect sizes. Although it is common for meta-analysts to set missing effect sizes to zero, Hedges and Becker (1986) strongly advise against it due to the reduction in effect size magnitude and the artificial suppression of variability. Therefore, if studies reported an effect but did not report quantitative information such as means and standard deviations or effect-sizes, the studies were excluded.

Overall, 109 articles met the defined criteria. Most of the excluded articles were removed due to not containing enough data to compute effect sizes (e.g., results were reported in aggregate and not by gender) and additional data could not be obtained from the authors. In four cases, authors reported multiple estimates of the same construct from the same sample (i.e. two self-estimates of verbal ability measures). In these cases, a mean effect size was computed to retain the independence of the estimates (Hedges & Becker, 1986). The included articles yielded 131 studies or independent samples and a total of 237 effect sizes. There were 82 effect sizes for overall academic self-concept, 74 effect sizes for verbal self-concept, and 81 effect sizes for mathematical self-concept.

2.2 Coding Studies

A standardized coding scheme based on the selection of moderator variables and required data for computing effect sizes was developed. Some variables were coded at the level of the individual study or independent sample, others were coded at the level of the individual effect size.

The following variables were coded at the level of the individual study or independent sample:

Author gender proportion. Four categories were used (1 = *majority male*; 2 = *majority female*; 3 = *equal number of male and female authors*; 4 = *undetermined*). Gender of each author was determined by the gender pronouns used in the author's publications or the bio on their institution's website.

Country of origin. Seven categories were used (1 = *Americas*; 2 = *United Kingdom*; 3 = *Europe [excluding the UK]*; 4 = *Australia or New Zealand*; 5 = *Africa [including the Middle East]*; 6 = *Asia*; 7 = *Multiple locations*). Country of origin was determined by where the data were collected. If the location of data collection was unknown, the location of the first author's affiliation was used. After all studies were coded, these were combined into three levels (Western, Non-Western, and Combined) due to multiple categories having less than 5 effect sizes.

Participant mean age. The mean age of the sample. If the mean age was not reported and the sample was described as "undergraduates" or "psychology students," a mean age of 20 was used. If a student sample was reported with a grade level but no age, 5 was added to the grade to approximate age.

Publication year. The year of the publication. If article included the year in which the data were collected, this year was used instead of publication date.

The following variables were coded at the level of the individual effect size:

Construct measured. Three categories were used (1 = *Academic self-concept*; 2 = *Self-estimates of ability [defined as single-item measures or an average of single-item measures of multiple intelligences]*; 3 = *Other/unknown*).

Assessment used. Eight categories were used (1 = *Academic Self-Concept Scale*; 2 = *Multiple Intelligences Inventory*; 3 = *Self-Concept of Ability Scale*; 4 = *Self-Description Questionnaire II*; 5 = *Self-Description Questionnaire III*; 6 = *Self-Rating Scale of College Ability*; 7 = *Other*; 8 = *Unspecified*).

Format of the assessment. Seven categories were used (1 = *Likert-like items*; 2 = *Labelled bell curve*; 3 = *Mean score of 100*; 4 = *Visual analogue scale*; 5 = *Percentile scale*; 6 = *Other*).

Number of items. Number of items used for assessing self-concept.

Specificity of items. Four categories were used (1 = *Domain [e.g., “Mathematics” or “English”]*; 2 = *Sub-domain [e.g., “Algebra” or “Vocabulary”]*; 3 = *Task level [e.g., “I enjoy creating puns” or “I am good at word puzzles”]*; 4 = *Unknown*).

Gender role orientation. Four categories were anticipated (1 = *Primarily Feminine*, 2 = *Primarily Masculine*, 3 = *Primarily Undifferentiated*, 4 = *Primarily Androgynous*). Although gender role orientation was a hypothesized moderator of

gender differences in overall academic, verbal, and mathematical self-concept, the search of the literature did not result in enough studies to be analyzed. Only four studies reported data and, of those studies, none of the findings were reported in a way that could be standardized for analysis. Despite contacting authors, no usable data were obtained.

2.3 Inter-rater Reliability

Twenty-five articles were randomly selected to be coded by the author and a second coder. These articles resulted in a total of thirty-two independent samples being coded for interrater reliability analysis. The average interrater reliability was satisfying at $\kappa = .94$. The lowest kappa statistic was for Overall Academic Self-Concept Format ($\kappa = .81$) and the highest was for Country of Origin, Overall Academic Self-Concept Construct Measured, and Mathematical Assessment Used ($\kappa = 1.0$). All discrepancies were discussed and, if necessary, additional clarification and definition was added to the coding guide. After the adjustments, the remaining studies were coded by the author.

2.4 Statistical Analyses

To approximate the magnitude of gender differences in academic self-concept, three separate meta-analyses were conducted on overall academic self-concept, verbal self-concept, and mathematical self-concept. Analyses were performed using formulas provided by Lipsey and Wilson (2001) and Hedges and Becker (1986).

The effect size, g , was calculated and defined as the mean for males minus the mean for females, divided by the pooled sample standard deviation (therefore, positive values indicated that males had higher self-ratings than females did). If means and standard

deviations were not available, effect sizes were computed using results of *F-tests* or *t-tests*. When insufficient information was available to compute effect sizes, the studies were eliminated as noted in the Method section. Due to the tendency for an upward bias in small samples, all effect sizes were weighted by an inverted variance resulting in the corrected effect size, d (Hedges & Becker, 1986; Lipsey & Wilson, 2001).

To avoid the distortion of the analyses due to outliers, Huffcut and Arthur's (1995) sample-adjusted meta-analytic deviancy statistic (SAMD) was used to determine studies for exclusion after the corrected d -values were obtained. This method of identifying outliers is more analytically precise due to the consideration of sample size. Based on the analysis, two outliers were removed for overall academic self-concept, four outliers were removed for verbal self-concept, and two outliers were removed for mathematical self-concept. For the interpretation of the effect sizes, Cohen's (1988) framework was used; d -values of $d = 0.20$ to $d = 0.49$, $d = 0.50$ to $d = 0.79$, and $d > 0.80$ will be characterized as small, medium, and large, respectively. Effect sizes of $d < 0.20$ will be considered negligible.

Initial analyses of homogeneity were computed for each self-concept measure using fixed-effect models. When the null hypothesis of homogeneity was rejected, mixed-effects models were used to evaluate the presence of moderator effects. In contrast to the fixed-effect model (which assumes that all variation is systematic), and the random effects model (which assumes that none of the variation is systematic), the mixed-effects model assumes that the moderation of between-subject variables is systematic, but that there is a remaining unmeasured random effect in the effect size distribution in addition to sampling error (Lipsey & Wilson, 2001). In mixed-effects models, a random-effects variance component

is computed based on the residual homogeneity after moderator effects have been considered. Then, inverse variance weights are recalculated using the random-effects variance. Lipsey and Wilson (2001) suggest that the mixed-effects model provides the lowest Type I error rate while still maximizing power for identifying moderator effects.

Additionally, rather than using the Hedges and Olkin (1985) method of adjusting the Type I error rate when using multiple Bonferroni comparisons across a meta-analysis, meta-regression was used to examine the relationship of the hypothesized moderators (publication year and number of items) on effect magnitude as recommended by Lipsey (2009) and Pigott (2012). This allows for the simultaneous consideration of the predictor variables rather than examining one moderator at a time through multiple ANOVAs which can lead to an enlarged Type I error rate. Single variable analyses were conducted for exploratory moderators using analogues to the analysis of variance for categorical variables and meta-regression models for continuous variables. If exploratory analyses resulted in significance for categorical variables, post hoc contrasts were performed using the Bonferroni method (Hedges & Olkin, 1985). For studies that included measures of both self-estimated intelligence constructs and objective measures of intelligence, analogues to the analysis of variation were used to compare the effect sizes. All analyses were conducted using the metafor R package by Viechtbauer (2010).

CHAPTER 3

Results

The meta-analyses of overall academic, verbal, and mathematical self-concepts revealed a similar pattern of sex differences across all constructs with men reporting higher self-estimates of ability and self-concepts than women reported on average. The largest difference was for mathematical self-concept ($d = 0.41$) and the smallest difference was for verbal self-concept ($d = 0.13$). For a summary of results, see Table 1.

Table 1. Summary of Meta-Analysis Results, as Moderated by Number of Items and Publication Year

Self-concept domain	R^2	Q_{model}	$B_{numitems}$	$B_{pubyear}$	k
Overall academic self-concept	.303	33.31	-.0109	.0049	80
Verbal self-concept	.023	1.39	.0106	.0018	70
Mathematical self-concept	.045	3.24	-.0269	-.0021	73

3.1 Overall Academic Self-Concept

3.1.1 Magnitude of gender differences

To test Hypothesis 1, the weighted mean effect size of gender differences in overall academic self-concept was calculated and evaluated using confidence intervals. The 80 effect sizes ranged from $d = -0.30$ to $d = 0.92$, with a weighted mean effect size of $d = 0.32$ and a 95% confidence interval of $0.29 < d < 0.34$. This indicates that, on average, males

estimated their overall academic ability approximately one third of a standard deviation higher than females. Though the effect size did not reach the predicted magnitude, the statistical significance of the mean difference supports Hypothesis 1.

Six cases reported women estimating their overall academic ability higher than men. Of the remaining 74 effect sizes favoring men, 14 cases reported negligible sex differences ($d < 0.20$), 42 cases reported small sex differences ($0.20 < d < 0.49$), 15 cases reported moderate sex differences ($0.5 < d < 0.79$), and three cases reported large sex differences ($0.8 < d$). The total participant sample size for these 80 effect sizes was 20,369, which corresponds to an average sample size of approximately 255 participants per study (the smallest sample was 40, and the largest sample size was 2,589). In order to determine whether the variability among the effect sizes was greater than would be expected from sampling error alone, homogeneity analysis was conducted using methods described by Lipsey and Wilson (2001). Results of the analysis revealed that the set of effect sizes was significantly heterogeneous, $Q_t(79) = 221.46, p < .001$, suggesting that the assumptions of the fixed-effect model should be rejected. Given that the Q statistic suggested that each effect size did not estimate a common population mean, the mixed-effects model was used to detect the presence of moderators.

3.1.2 Hypothesized moderators

As previously discussed, the mixed-effects model assumes that the effect size variation can be divided into two components: systematic relationships between characteristics of the studies and the effect sizes, and random study-level variance (Lipsey & Wilson, 2001). This model enables the researcher to evaluate the influence of between-

study moderator variables and both unmeasured random variance and subject-level sampling error.

Weighted least squares regression using the method of moments based estimate of the random effects variance component was used to evaluate moderator hypotheses. Hypothesis 5 anticipated that the publication year of the study would moderate the magnitude of effect sizes such that more recent studies would have smaller effect sizes and older studies would have larger effect sizes. Hypothesis 6 predicted that as the number of items on a measure increased, the effect size for gender differences in self-concept would decrease. Therefore, publication year and number of items were entered into a mixed-effects weighted least squares regression as predictor variables. The model significantly predicted heterogeneity among the effects, $Q_{model}(2) = 33.31, p < .0001$. This suggests that the relationships captured by the model are stronger than we would expect by chance (Borenstein, Hedges, Higgins, & Rothstein, 2009). As predicted, the number of items used to assess overall academic self-concept moderated gender differences in the expected direction ($\beta = -.01, p < .0001$). The date of publication also significantly moderated gender differences, however, it did so in the opposite direction as was anticipated ($\beta = .01, p < .0001$). Thus, Hypothesis 5 was supported for overall academic self-concept, but Hypothesis 6 was not.

3.2 Verbal Self-Concept

3.2.1 Magnitude of gender differences

The weighted mean effect size, d , of the gender differences in verbal self-concept was .13, which indicates a minimal difference favoring men as predicted by Hypothesis 2. The 70 effect sizes ranged from $d = -0.42$ to $d = 0.56$ and had a 95% confidence interval

of $0.10 < d < 0.16$. In 13 studies, effect sizes ranged from -0.30 to -0.01 favoring females. Among the samples with effect sizes favoring males, 34 cases reported negligible sex differences ($0.00 < d < 0.20$), 21 cases reported small sex differences ($0.20 < d < 0.49$), and two cases reported moderate effect sizes ($0.50 < d < 0.79$). There were no effect sizes for gender differences in verbal self-concept that reached large differences. The total participant sample was 17,508, which averages to approximately 250 participants per sample.

To determine the appropriateness of moving forward with moderator analysis, a test of homogeneity was conducted. The homogeneity statistic revealed substantial variability within the sample of verbal self-concept studies, $Q(69) = 202.34, p < .001$, leading to the rejection of fixed-effect assumptions. Therefore, the mixed-effects model was applied to the data.

3.2.2 Hypothesized moderators

Hypotheses 5 and 6 were also evaluated for verbal self-concept using the weighted least squares regression model containing publication year and number of items as predictors. However, unlike overall academic self-concept, this model fit the verbal self-concept data poorly and did not significantly predict heterogeneity among the effect sizes for verbal self-concept ($Q_{model}(2) = 1.39, p = .49$). Therefore, the individual moderators, publication year and number of items, were not investigated.

3.3 Mathematical Self-Concept

3.3.1 Magnitude of gender differences

The third and final meta-analysis also resulted in a significant weighted mean effect size favoring men ($d = 0.42$). The 79 effect sizes for mathematical self-concept ranged from $d = -0.16$ to $d = 0.92$ and had a 95% confidence interval of $0.39 < d < 0.44$. In all but three studies, men's estimations of their mathematical intelligence were higher than women's estimations. Only eight cases reported negligible sex differences ($d < 0.20$), 35 cases reported small sex differences ($0.20 < d < 0.49$), 28 cases reported moderate sex differences ($0.50 < d < 0.79$), and 5 cases reported large differences ($0.80 < d$). An average of 254 participants per sample resulted in a total participant sample of 20,057.

As with both overall academic and verbal self-concepts, the homogeneity statistic for mathematical self-concept effect sizes revealed substantial variability within the sample of studies, $Q_I(78) = 234.23, p < .001$. This led to the rejection of fixed-effect assumptions and continuance with the mixed-effects model.

3.3.2 Hypothesized moderators

Using the weighted least squares regression model, publication year and number of items were evaluated as moderators. As with verbal self-concept, this model fit the mathematical self-concept data poorly and did not predict heterogeneity among the effect sizes, $Q_{model}(2) = 3.24, p = .198$. Hypotheses 5 and 6 were not supported.

3.4 Exploratory Analyses

3.4.1 Age, author gender, and country of origin

Exploratory analyses of age, author gender ratio, and country of origin were conducted using meta-regression or analogues to ANOVA for overall academic self-concept, verbal self-concept, and mathematical self-concept. These analyses resulted in

non-significant findings; relevant statistics can be found in Tables 2 and 3. It is interesting to note that when run as a mixed-effects model, none of the exploratory moderators were found to be significant. However, when replicating fixed-effect methods used by Furnham and Syzmanowicz (2011), the results for many of the moderators were significant.

3.4.2 Objectively measured intelligence

Effect sizes for gender differences in objective measures of overall academic, verbal, and mathematical intelligence were compared to the related self-estimate measure effect sizes using analogues to ANOVA. For each measure of intelligence, fewer than 10 samples reported gender differences in both objective and self-concept measures to be used in analysis.

Overall academic intelligence had the largest k , with six studies reporting the necessary statistics for analysis. The weighted mean effect size for gender differences in overall academic self-concept was $d = 0.37$ (favoring men) and the weighted mean effect size for gender differences in objective intelligence was $d = -0.22$, (favoring women). The finding was insignificant, $Q_{between}(1) = 2.05, p = .153$. For verbal intelligence, only three studies were available. The resulting d -value for gender differences in objective verbal intelligence was 0.19 and the d -value for verbal self-concept favored women ($d = -0.13$). The result of the analogue to ANOVA was non-significant, $Q_{between}(1) = 2.73, p = .099$. The final comparison was between objective measures of mathematical intelligence and mathematical self-concept ($k = 4$). For mathematical intelligence, men scored higher on both the self-concept measures ($d = 0.43$) and objective measures ($d = 0.98$). This contrast was also insignificant, $Q_{between}(1) = 1.75, p = .186$.

CHAPTER 4

Discussion

This study used meta-analytic techniques to investigate the effect sizes of gender differences in overall, mathematical, and verbal self-concept. The results support the hypotheses of mean effect size differences favoring men in all three domains. However, many of the moderator analyses resulted in null findings.

Males reported significantly higher ratings of self-concept than females for overall academic ($d = 0.32$), verbal ($d = 0.13$), and mathematical ($d = 0.41$) self-concepts. Although not as large in magnitude as expected, gender differences in overall academic and mathematical self-concepts favored men as anticipated (Hypotheses 1 and 2). Hypothesis 3 was also supported with men scoring significantly higher on measures of verbal self-concept compared to women, though the effect size was negligible as predicted.

4.1 Hypothesized Moderator Variables

Given the results of the test for heterogeneity, analyses for hypothesized and exploratory moderator variables were conducted for overall academic, verbal, and mathematical self-concept. Although many of the hypotheses for moderators were not supported or were unable to be explored, these outcomes suggest possibilities for future research and highlight potential methodological weaknesses in previous research. In addition to the exploratory analyses, my hypothesized moderators were sex-type, publication year, and number of items.

Based on findings that sex-typed individuals are more likely to adhere to traditional gender roles, I anticipated that comparisons of gender-typed individuals would result in

larger disparities between men and women than comparisons of non-gender-typed persons. Gender-type, as measured by Bem's Sex-Role Inventory, could potentially influence the magnitude of gender differences for multiple reasons. It has been proposed that certain domains of intelligence are stereotypically gendered, with math being traditionally seen as a male domain and verbal ability being typed as feminine (Guimond, Chatard, Martinot, Crisp, & Redersdorff, 2006; Nosek et al., 2009). Additionally, Furnham, Hosoe, and Li-Pang Tang (2001) posited that, at least in Western cultures, general intelligence is judged by weighting male-normed components of intelligence (mathematical ability and spatial ability) more heavily than stereotypically female aspects of intelligence (verbal ability). This focus on math and spatial ability could lead to the inflation of gender differences on overall academic self-concept measures. If intelligence is seen as male-normative and men are expected to score higher than women, gender-typed individuals would be more likely to adhere to these stereotypes and under- or overestimate their cognitive ability accordingly. This moderator would lead to an increase in gender differences in overall academic and mathematical self-concept, and reduce the magnitude of differences in verbal self-concept.

Ideally, the influence of gender-type on the magnitude of sex differences on measures of self-concept would be investigated by comparing the difference between sex-typed males and females to the difference between androgynous or undifferentiated males and females. However, the already limited number of studies exploring gender-type and self-concept did not structure their analyses in ways that allowed for standardization for a meta-analysis. Due to this limitation, I was unable to explore the relationship between

sex-type and gender differences in self-concept. Future directions for research in this area and the potential implications are discussed in later sections.

The influence of publication year and number of items on the magnitude of gender differences varied across overall academic, verbal, and mathematical self-concepts. For overall academic self-concept, the regression model using publication year and number of items as predictor variables explained significant variability across the effect sizes. Both moderators were significant, though publication year was significant in the opposite directionality as anticipated.

It was hypothesized that as the number of items on a measure of self-concept increased, the magnitude of gender differences would decrease due to the ambiguity of the single item measures. Though the distinction between self-assessed intelligence and academic self-concept continues to be debated (Freund & Kasten, 2012; Peterson & Whiteman, 2007), one consistent difference is how the constructs are most frequently measured. Due to self-concept being comprised of cognition, affect, and behavior (Byrne, 1996), measures of academic self-concept are generally composed of multiple items (e.g., the Self-Description Questionnaire III [Marsh, 1992] or the Academic Self Concept Scale [Reynolds, Ramirez, Magriña, & Allen, 1980]). Conversely, measures of self-assessed intelligence or self-estimates of intelligence commonly use Furnham's (2000) method of requesting a single estimate for each domain using an image of a normal distribution with 100 as its mean.

The importance of the number of items on a measure is two-fold. First, the tendency for individuals to assess themselves more positively than they assess others (Brown, 2012)

is exacerbated when the trait being evaluated is measured via ambiguous items (Dunning, Meyerowitz, & Holzberg, 1989; Felson, 1981). When the criteria for making an evaluation of oneself is vague, it allows participants to use disparate criteria and a broader range of experiences on which to base their assessments. Secondly, the difference in how men and women interpret information related to failures and successes could also be impacted by the number of items on a measure. When recalling self-performance and commentary from others, women recall both positive and negative feedback while males retain only positive experiences (Roberts, 1991). This masculine trait of positive recollection leads to men inflating their self-concept on all measures of intelligence and results in women having seemingly more humility (Syzmanowiz & Furnham, 2013). The likelihood of men to rely primarily on positive experiences is further exacerbated by items that do not provide parameters or specific criteria to guide the self-estimations (Thomas, 1991; von Stumm, 2014). For overall academic self-concept the number of items hypothesis was supported and was a significant moderator of mean effect size differences between men and women.

The second moderator that was analyzed was publication year. It has been argued that gender differences across a variety of domains are declining due to women's roles becoming more comparable to men's roles (Byrne, Miller, & Schafer, 1999; Hyde, 2005). Therefore, it was hypothesized that more recent studies would have smaller gender differences than older studies. This hypothesis was not supported and, for overall academic self-concept, the result was significant in the opposite direction such that newer articles had larger effect sizes. Though this finding was unexpected, after considering the magnitude of gender differences in self-concept when compared to gender differences in objective ability, it is possibly explained by the type of construct being measured. Recent

studies of the gender similarities hypothesis found gender differences to be small to nonexistent overall and for publication year to significantly moderate the magnitude of the effect sizes such that newer studies have small effect sizes than older studies (Hyde, 2005). However, the literature on the gender similarities hypothesis primarily focuses on objective measures of traits such as cognitive abilities and psychopathology. When examining constructs dealing with self-perception or self-reported traits, the similarity hypothesis might not hold true. Zell, Krizan, and Teeter (2015) found that the largest d -value in their metasynthesis of gender differences was for masculine versus feminine traits ($d = .73$). Gender stereotypes continue to be salient in many cultures and research shows that children and adults alike tend to conform to societal expectations of gender norms (Eagly & Wood, 2013). It is possible that, although gender differences in objective intelligence are diminishing over time, the long-held belief that males perform better than females on intelligence measures continues to be perpetuated and is leading to discrepant academic self-concepts between men and women.

For both mathematical and verbal self-concepts, the regression models using publication year and number of items did not account for a significant amount of variability across the effect sizes. Although the moderator hypotheses were not supported, the results still warrant additional investigation. For number of items, it is possible that the level of ambiguity across the various measures of self-concept were not differentiated enough to influence the magnitude of gender differences. For overall academic self-concept, the number of items on a measure ranged from one to 40, but for mathematical and verbal self-concept, the number of items only ranged from one to 10. It could be that measures with more items did not provide enough specification to lessen the impact of men's likelihood

for positive recollection. In the future, it would be interesting to learn what experiences individuals are using to determine their self-ratings and see how the level of ambiguity of a measure might impact those determinations. The reasons for publication year not being a significant moderator for mathematical and verbal self-concepts could potentially be the same as those outlined for overall academic self-concept.

4.2 Exploratory Moderator Variables

Age, author gender ratio, and country of origin were investigated as exploratory moderator variables for overall academic, verbal, and mathematical self-concepts. All findings were null. Though there was theoretical justification for the exploration of each of the moderators, the primary support from previous research came from Syzmanowicz and Furnham's (2011) meta-analysis. It is possible that the current meta-analyses did not replicate the 2011 study due to differences in the samples identified for analysis, but it seems more likely that the inconsistency in findings came from the use of more stringent methodology. Syzmanowicz and Furnham (2011) used fixed-effect models for all analyses of gender differences in self-estimated intelligences. Though fixed-effect models were once the norm for meta-analyses, they have been criticized and more rigorous methods have been developed (Hedges & Vevea, 1998; Lipsey & Wilson, 2001). The major criticism of fixed-effect models is that the method assumes all variability among the effect sizes is wholly systematic and can be completely accounted for by the moderators. This assumption is generally untenable and leads to an inflation of the Type I error rate (Borenstein et al., 2009; Hedges & Vevea, 1998; Lipsey & Wilson, 2001). It is possible that my findings did not replicate those of Syzmanowicz and Furnham (2011) due to the inflation of Type I error rates that come with fixed-effect models. In fact, when I conducted

my exploratory moderator analyses using the same methods as the 2011 meta-analysis, many of the previously null outcomes were significant (see Tables 2 and 3 for comparisons).

4.3 Objective Intelligence

Despite finding significant gender differences favoring men on all three measures of self-concept, it is unlikely that these gender differences reflect inequalities on objective intelligence measures. In an attempt to compare the magnitude of gender differences between self-estimated intelligence and measured intelligence, the weighted mean effect sizes were evaluated. However, due to the very limited number of studies reporting gender differences for both self-concept and psychometrically evaluated intelligence, the results of this study are not likely to generalize.

All comparisons between self-concept and objective intelligence measures were likely insignificant due to a lack of power. As explained by Hedges and Piggott (2004), moderator analyses are conceptually the same as interaction analyses and therefore have less power than tests for main effects in the same designs. Therefore, given the very small *ks* for each measure, it is unsurprising that the tests were not sensitive enough to detect differences. For overall intelligence, only six studies reported both self-report measures and objective measures of ability; for verbal intelligence, only three studies were available. Additionally, the limited number of available studies that reported findings in a comparable way makes it unlikely that the analyses would be generalizable.

The null results for mathematical ability were particularly unexpected when we examine the weighted mean effect sizes for objective intelligence. Though males did, at

Table 2. Summary of Meta-Analysis Results, as Moderated by Categorical Exploratory Sample Characteristics

Sample Characteristics	Mixed-Effects				Fixed Effects			
	<i>Q</i>	df	<i>p</i>	Mean ES	<i>Q</i>	df	<i>p</i>	Mean ES
Overall academic self-concept								
Author gender ratio								
Total	74.24	79	.63	.33	221.46	79	<.001	.31
Between	0.59	2	.74		1.19	2	.55	
Within	73.65	77	.58		220.26	77	<.001	
More male authors	45.24	49	.63	.34	137.93	49	<.001	.31
More female authors	18.82	17	.34	.35	50.54	17	<.001	.35
Equal authors	9.59	11	.57	.28	31.79	11	<.001	.31
Nationality								
Total	77.06	79	.54	.33	221.46	79	<.001	.31
Between	2.41	2	.29		8.95	2	.01	
Within	74.65	77	.55		212.51	77	<.001	
Western Samples	63.37	63	.46	.32	178.55	63	<.001	.30
Non-Western Samples	10.43	10	.40	.36	31.25	10	<.001	.34
Mixed Samples	0.85	4	.93	.47	2.69	4	.61	.47
Verbal self-concept								
Author gender ratio								
Total	54.47	69	.89	.10	202.34	69	<.001	.13
Between	0.67	2	.71		0.42	2	.81	
Within	53.79	67	.87		201.91	67	<.001	
More male authors	14.15	35	.99	.11	42.79	35	.17	.13
More female authors	25.19	23	.34	.08	121.91	23	<.001	.13
Equal authors	14.45	9	.11	.13	31.22	9	<.001	.10
Nationality								
Total	51.13	69	.50	.13	202.34	69	<.001	.13
Between	1.37	2	.94		5.46	2	.49	
Within	49.76	67	.94		196.88	67	<.001	
Western Samples	27.63	47	.98	.13	74.54	47	<.01	.10
Non-Western Samples	19.51	16	.24	.11	111.00	16	<.001	.15
Mixed Samples	2.62	4	.62	.24	11.34	4	.02	.22
Mathematical self-concept								
Author gender ratio								
Total	78.18	78	.25	.43	234.23	78	<.001	.41
Between	2.79	2	.49		1.29	2	.52	
Within	75.41	76	.47		232.93	76	<.001	
More male authors	32.22	37	.69	.48	75.49	37	<.001	.42
More female authors	21.79	20	.35	.38	70.42	20	<.001	.44
Equal authors	21.39	19	.32	.39	87.02	19	<.001	.39
Nationality								
Total	78.59	78	.36	.43	234.23	78	<.001	.41
Between	2.06	2	.46		4.69	2	.09	
Within	76.53	76	.46		229.53	76	<.001	
Western Samples	56.05	54	.39	.46	165.53	54	<.001	.43
Non-Western Samples	15.11	17	.59	.39	51.43	17	<.001	.37
Mixed Samples	5.37	5	.37	.10	12.57	5	.03	.35

Note. *P*-values < .05 are in boldface. Mean ES = Weighted mean effect size.

Note. *P*-values < .05 are in boldface

Table 3. Summary of Meta-Analysis Results, as Moderated by Age

Self-concept domain	Mixed-Effects					Fixed Effects				
	R^2	Q_{model}	β	p	k	R^2	Q_{model}	β	p	k
Overall academic self-concept	.0356	2.79	.0060	.0943	79	.0535	11.82	.0073	<.001	79
Verbal self-concept	.0649	3.61	.0051	.0571	69	.0227	4.53	.0043	.033	69
Mathematical self-concept	.0416	3.31	.0063	.0690	77	.0000	.0015	.0001	.9689	77

one time, outperform females on mathematical abilities (Maccoby & Jacklin, 1974), this stereotypical “truth” has been challenged by multiple reviews. In fact, recent meta-analytic results show a decline in the magnitude of gender differences on measures of mathematical performance since the initial evaluation of the literature in 1974 (Hyde, 2014). Recent findings suggest that the difference overall is minimal (sometimes even nonexistent, $d = -.069$; Voyer & Voyer, 2014) and that larger sex differences in mathematics only occur in specific instances (Lindberg, et al. 2010). However, the weighted mean average for mathematical ability of the four studies identified for this analysis was $d = .98$ (with a range of $d = 0.07$ to $d = 2.34$), far exceeding the results of any meta-analytic review in the last thirty years. This exceptionally large value was possibly driven by extreme effect sizes due to having such a limited sample of studies in the analysis.

4.4 Summary

Many of the results were null or ran contrary to expectation, but potentially informative results were observed for mean effect size differences and for moderators of overall intelligence. In all three self-concept domains, men estimated their own intelligence to be significantly higher than females estimated theirs. Though the effect sizes were smaller than anticipated for overall academic and mathematical self-concept, the

significance suggests that there are notable mean gender differences on these measures. Additionally, despite verbal ability being a stereotypically female-oriented intelligence and women typically scoring higher on objective measures, results were still significant favoring men (though the effect size was negligible). Finally, the mixed results from both hypothesized and exploratory moderators suggests avenues of future research.

4.5 Implications

Overall, the lack of available data to evaluate the impact of sex-type on academic self-concept or to compare self-concept to objective measures makes it difficult to assess any true implications. Findings suggest that men have higher self-concepts for overall academic, verbal, and mathematical domains than women, but this study was unable to adequately address the underlying influences on the magnitude of those differences. In this regard, the results are disappointing. It is commonly posited that gender differences in self-concept are due to female humility and male hubris (Furnham, 2001), but without the appropriate data being available, this study was unable to test this hypothesis. To thoroughly investigate the relationships between sex-role adherence, academic self-concept, and objective intelligence, further research is required.

4.6 Limitations

The first and largest limitation of this study was the limited availability of relevant data. Though I was able to identify a plethora of pertinent studies, a large portion of the articles did not report the necessary statistics for computing effect sizes and authors did not respond to requests for additional data. Additionally, one of my primary moderators of interest, Bem's Sex Role Orientation, was unable to be explored due to the extremely small

number of articles investigating the relationship between sex-role adherence and academic self-concept.

As with all meta-analyses, a possible limitation is the ‘file-drawer’ problem (Rosenthal, 1979). There is a bias towards publishing statistically significant results (Lipsey & Wilson, 2001), and research shows that published articles have a larger mean effect size than unpublished articles (Lipsey & Wilson, 1993). However, because published articles are more accessible, meta-analyses often overrepresent significant findings, causing an upward bias in the results. Though authors were contacted to obtain unpublished works, I received no additional studies and was not able to lessen the file-drawer problem through the inclusion of unpublished data.

4.7 Conclusion

This study represents an attempt to examine gender differences in multiple domains of self-concept using meta-analytic methods. Though significant weighted mean effect size differences were found for overall academic, verbal, and mathematical self-concepts, most hypotheses regarding the influence of moderating variables were not supported. Findings of significant gender differences in self-concept in all three domains are an interesting contrast to recent findings of limited gender differences in objective intelligence. Additionally, there was a notable lack of articles containing data for Bem’s Sex Role Inventory, as well as a lack of articles reporting both self-concept and objective intelligence measures. These limitations left me unable to adequately address the underlying questions of why gender differences exist in self-concept and whether the gender differences in self-concept reflect actual differences on objective measures. Future research should investigate the directionality of any discrepancies between self-concept and objective measures to

better understand how traditional gender stereotypes might be influencing academic self-concept.

APPENDIX

Articles Used for Meta-Analyses.

- Ackerman, P. L., & Ellingsen, V. J. (2014). Vocabulary overclaiming – A complete approach: Ability, personality, self-concept correlates, and gender differences. *Intelligence*, 46, 216-227.
- Ackerman, P. L., Shapiro, S., & Beier, M. E. (2011). Subjective estimates of job performance after job preview: Determinants of anticipated learning curves. *Journal of Vocational Behavior*, 78(1), 31-48.
- Ackerman, P. L., & Wolman, S. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, 13, 57-78.
- Awad, G. H. (2007). The role of racial identity, academic self-concept, and self-esteem in the prediction of academic outcomes for African American students. *Journal of Black Psychology*, 33(2), 188-207.
- Bailey, K. G., & Lazar, J. (1976). Accuracy of self-ratings of intelligence as a function of sex and level of ability in college students. *Journal of Genetic Psychology*, 129, 279-290.
- Bailey, K. G., & Minor, S. W. (1976). Self-image congruence in freshmen, seniors, and graduate students. *Journal of Genetic Psychology*, 129, 301-309.
- Bailey, R. C., & Bailey, K. G. (1971). Perceived ability in male and female college students. *Perceptual and Motor Skills*, 32(1), 293-294.
- Bailey, R. C., & Mettetal, G. W. (1977). Perceived intelligence in married partners. *Social Behavior and Personality*, 5(1), 137-141.
- Bandalos, D. L., Yates, K., & Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *Journal of Educational Psychology*, 87(4), 611-623.
- Barnard, L., & Olivarez, A. (2007). Self-estimates of multiple, g factor, and school-valued intelligences. *North American Journal of Psychology*, 9(3), 501-510.
- Beier, M. E., & Ackerman, P. L. (2001). Current-events knowledge in adults: An investigation of age, intelligence, and nonability determinants. *Psychology and Aging*, 16(4), 615-628.
- Beloff, H. (1992). Mother, father, and me: Our IQ. *The Psychology*, 5, 309-311.
- Bennett, M. (1996). Men's and women's self-estimates of intelligence, *The Journal of Psychology*, 136(3), 411-412.

- Bennett, M. (1997). Self-estimates of ability in men and women. *Journal of Social Psychology, 137*(4), 540-541.
- Bennett, M. (2000). Self-estimates and population estimates of ability in men and women. *Australian Journal of Psychology, 52*, 23-28.
- Bonnot, V., & Croizet, J. (2007). Stereotype internalization, math perceptions, and occupational choices of women with counter-stereotypical university majors. *Swiss Journal of Psychology, 66*, 169-178.
- Bonnot, V., & Jost, J. T. (2013). Divergent effects of system justification salience on the academic self-assessments of men and women. *Group Processes & Intergroup Relations, 17*(4), 453-464.
- Caskie, G. I. L., Sutton, M. C., & Eckhardt, A. G. (2014). Accuracy of self-reported college GPA: Gender-moderated differences by achievement level and academic self-efficacy. *Journal of College Student Development, 55*(4), 385-390.
- Chamorro-Premuzic, T., & Furnham, A. (2006). Personality and self-assessed intelligence: Can gender and personality distort self-assessed intelligence? *Educational Research and Reviews, 1*(7), 227.
- Cokley, K. (2000). An investigation of academic self-concept and its relationship to academic achievement in African American college students. *Journal of Black Psychology, 26*(2), 148-164.
- Cokley, K. (2001). Gender differences among African American students in the impact of racial identity on academic psychosocial development. *Journal of College Student Development, 45*(5), 480-487.
- Cokley, K. (2002). Ethnicity, gender and academic self-concept: A preliminary examination of academic disidentification and implications for psychologists. *Cultural Diversity and Ethnic Minority Psychology, 8*(4), 378-388.
- Cokley, K., & Moore, P. (2007). Moderating and mediating effects of gender and psychological disengagement on the academic achievement of African American college students. *Journal of Black Psychology, 33*(2), 169-187.
- Cokley, K., Awad, G., Smith, L., Jackson, S., Awosogba, O., Hurst, A., Stone, S., Blondeau, L. & Roberts, D. (2015). The roles of gender stigma consciousness, impostor phenomenon and academic self-concept in the academic outcomes of women and men. *Sex Roles, 73*(9-10), 414-426.
- Cokley, K., McClain, S., Jones, M., & Johnson, S. (2012). A preliminary investigation of academic disidentification, racial identity, and academic achievement among African American adolescents. *The High School Journal, 95*(2), 54-68.

- Cruise, S. M., Alan Lewis, C., & Mc Guckin, C. (2009). Examination of importance ratings and self-estimates in ten domains of intelligence: Evidence among a sample of UK and Irish university students. *The Irish Journal of Psychology*, 30(3-4), 171-183.
- Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly*, 37(3), 293-309.
- Furnham, A. (2000). Parents' estimates of their own and their children's multiple intelligences. *British Journal of Developmental Psychology*, 18, 583-594.
- Furnham, A. (2004). Are lay people lumpers or splitters? The factor structure of, and sex differences related to, self-rated and other-rated abilities. *Learning and Individual Differences*, 14(3), 153-168.
- Furnham, A. (2005). Self-estimated intelligence, psychometric intelligence and personality. *Psychologia*, 48(3), 182-192.
- Furnham, A. (2009). Sex, IQ, and emotional intelligence. *Psychological Reports*, 105, 1092-1094.
- Furnham, A., & Akande, A. (2004). African parents' estimates of their own and their children's multiple intelligences. *Current Psychology*, 22, 281-294.
- Furnham, A., Arteche, A., Chamorro-Premuzic, T., Keser, A., & Swami, V. (2009). Self- and other-estimates of multiple abilities in Britain and Turkey: A cross-cultural comparison of subjective ratings of intelligence. *International Journal of Psychology*, 44(6), 434-442.
- Furnham, A., & Baguma, P. (1999). Self-Estimates of Intelligence: A Cross-Cultural Study from Three Continents. *North American Journal of Psychology*, 1(1), 69-78.
- Furnham, A., & Buchanan, T. (2005). Personality, gender, and self-perceived intelligence. *Personality and Individual Differences*, 39, 543-555.
- Furnham, A., & Budhani, S. (2002). Sex differences in the estimated intelligence of school children. *European Journal of Personality*, 16(3), 201-219.
- Furnham, A., & Bunclark, K. (2006). Sex differences in parents' estimations of their own and their children's intelligence. *Intelligence*, 34, 1-14. doi:10.1016/j.intell.2005.05.005
- Furnham, A., Callahan, I., & Akande, D. (2004). Self-estimates of intelligence: A study in two African countries. *The Journal of psychology*, 138(3), 265-285.
- Furnham, A., & Chamorro-Premuzic, T. (2005). Estimating one's own and one's relatives' multiple intelligence: A study from Argentina. *The Spanish Journal of Psychology*, 8(1), 12-20.

- Furnham, A., Clark, K., & Bailey, K. (1999). Sex differences in estimates of multiple intelligences. *European Journal of Personality*, 13(4), 247-259.
- Furnham, A., & Crawshaw, J. (2002). School children's estimation of their own and their parents WISC scores. *Psychologia*, 45(1), 54-67.
- Furnham, A., Crawshaw, J., & Rawles, R. (2006). Sex differences in self-estimates on two validated IQ test subscale scores. *Journal of Applied Social Psychology*, 36(2), 417-440.
- Furnham, A., Dixon, D., Harrison, T., Rasmussen, S., & O'Connor, R. (2000). Sex, social class, and estimating IQ: middle-class male subjects rate themselves most intelligent. *Psychological reports*, 87(3), 753-758.
- Furnham, A., & Fong, G. (2000). Self-estimated and psychometrically measured intelligence: A cross-cultural and sex differences study of British and Singaporean students. *North American Journal of Psychology*.
- Furnham, A., Fong, G., & Martin, N. (1999). Sex and cross-cultural differences in the estimated multi-faceted intelligence quotient score for self, parents and siblings. *Personality and Individual Differences*, 26(6), 1025-1034.
- Furnham, A., & Fukumoto, S. (2008). Japanese parents' estimates of their own and their children's multiple intelligences: Cultural modesty and moderate differentiation. *Japanese Psychological Research*, 50, 63-76.
- Furnham, A., Hosoe, T., & Li-Ping Tang, T. (2002). Male hubris and female humility? A cross-cultural study of ratings of self, parental, and sibling multiple intelligence in America, Britain, and Japan. *Intelligence*, 30, 101-115.
- Furnham, A., Kidwai, A., & Thomas, C. (2001). Personality, psychometric intelligence, and self-estimated intelligence. *Journal of Social Behavior and Personality*, 16, 97-114.
- Furnham, A., & Mansi, A. (2014). The self-assessment of the Cattell–Horn–Carroll broad stratum abilities. *Learning and Individual Differences*, 32, 233-237.
- Furnham, A., Mkhize, N., & Mndaweni, T. (2004). Indian and isiZulu-speaking South African parents' estimates of their own and their children's intelligence. *South African Journal of Psychology*, 34(3), 364-385.
- Furnham, A., & Mottabu, R. (2004). Sex and Culture Differences in the Estimates of General and Multiple Intelligence: A Study Comparing British and Egyptian Students. *Individual Differences Research*, 2(2).
- Furnham, A., Moutafi, J., & Chamorro-Premuzic, T. (2005). Personality and intelligence: Gender, the big five, self-estimated and psychometric intelligence. *International Journal of Selection and Assessment*, 13(1), 11-24.

- Furnham, A., & Petrides, K. V. (2004). Parental estimates of five types of intelligence. *Australian Journal of Psychology*, 56(1), 10-17.
- Furnham, A., Rakow, T., & Mak, T. (2002). The determinants of parents' beliefs about the intelligence of their children: A study from Hong Kong. *International Journal of Psychology*, 37(6), 343-352.
- Furnham, A., Rakow, T., Sarmany-Schuller, I., & De Fruyt, F. (1999). European differences in self-perceived multiple intelligences. *European Psychologist*, 4, 131-140.
- Furnham A., & Rawles, R. (1995). Sex differences in the estimation of intelligence. *Journal of Social Behavior and Personality*, 10(3), 741-748.
- Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of genetic psychology*, 163(1), 24-39.
- Furnham, A., Shahidi, S., & Baluch, B. (2002). Sex and culture differences in perceptions of estimated multiple intelligence for self and family: A British-Iranian comparison. *Journal of Cross-Cultural Psychology*, 33, 270-285.
- Furnham, A., & Shagabutdinova, K. (2012). Sex differences in estimating multiple intelligences in self and others: A replication in Russia. *International Journal of Psychology*, 47(6), 448-459.
- Furnham, A., Tang, T. L. P., Lester, D., O'connor, R., & Montgomery, R. (2002). Estimates of ten multiple intelligences: Sex and national differences in the perception of oneself and famous people. *European Psychologist*, 7(4), 245.
- Furnham, A., & Thomas, C. (2004). Parents' gender and personality and estimates of their own and their children's intelligence. *Personality and Individual Differences*, 37(5), 887-903.
- Furnham, A., Tu, B. L., & Swami, V. (2012). Cross-cultural differences in self-assessed intelligence: a comparison of British and Chinese undergraduates. *Psychologia*, 55(1), 21-27.
- Furnham, A., & Valgeirsson, H. (2007). Parents' estimations of their own intelligence and that of their children: A comparison between English and Icelandic parents. *Scandinavian Journal of Psychology*, 48(4), 289-298.
- Furnham, A., von Stumm, S., Makendrayogam, A., & Chamorro-Premuzic, T. (2009). A Taxonomy of Self-Estimated Human Performance: The General Factor. *Journal of Individual Differences*, 30(4), 188-193.
- Furnham, A., & Ward, C. (2001). Sex differences, test experience and the self-estimation of multiple intelligences. *New Zealand Journal of Psychology*, 30(2), 52.

- Furnham, A., & Wu, J. (2008). Gender differences in estimates of one's own and parental intelligence in China. *Individual Differences Research*, 6(1), 1-12.
- Furnham, A., & Wu, J. (2014). The little emperor: Chinese parents' assessment of their own, their partner's and their only child's intelligence. *High Ability Studies*, 25(2), 121-141.
- Furnham, A., Wytykowska, A., & Petrides, K. V. (2005). Estimates of multiple intelligences: A study in Poland. *European Psychologist*, 10(1), 51-59.
- Furnham, A., Zhang, J., & Chamorro-Premuzic, T. (2005). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, cognition and personality*, 25(2), 119-145.
- Gati, I., Fishman-Nadav, Y., & Shiloh, S. (2006). The relations between preferences for using abilities, self-estimated abilities, and measured abilities among career counseling clients. *Journal of Vocational Behavior*, 68(1), 24-38.
- Hamid, P. N., & Lok, D. (1995). Gender stereotyping in estimates of intelligence in Chinese students. *The Journal of social psychology*, 135(3), 407-409.
- Holling, H., & Preckel, F. (2005). Self-estimates of intelligence—methodological approaches and gender differences. *Personality and Individual Differences*, 38(3), 503-517.
- Jordan, T. J., Grallo, R., Deutsch, M., & Deutsch, C. P. (1985). Long-term effects of early enrichment: A 20-year perspective on persistence and change. *American Journal of Community Psychology*, 13(4), 393-415.
- Kaufman, J. C. (2012). Self estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences*, 22, 118-222.
- Kirkcaldy, B., Noack, P., Furnham, A., & Siefen, G. (2007). Parental estimates of their own and their children's intelligence. *European Psychologist*, 12(3), 173-180.
- Kudrna, L., Furnham, A., & Swami, V. (2010). The influence of social class salience on self-assessed intelligence. *Social Behavior and Personality: an international journal*, 38(6), 859-864.
- Lips, H. M. (2004). The gender gap in possible selves: Divergence of academic self-views among high school and university students. *Sex Roles*, 50(5), 357-371.
- Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European journal of psychology of education*, 27(4), 483-498.
- Mucherah, W., Finch, W. H., & Keaikitse, S. (2012). Differential Bundle Functioning Analysis of the Self-Description Questionnaire Self-Concept Scale for Kenyan

- Female and Male Students Using the MIMIC Model. *International Journal of Testing*, 12(1), 78-99.
- Nasser, R., & Singhal, S. (2006). How youth in India and Lebanon rate their intelligence. *Journal of Social Sciences*, 2, 93-99.
- Neto, F., & Furnham, A. (2006). Gender differences in self-rated and partner-rated multiple intelligences: A Portuguese replication. *The Journal of Psychology*, 140(6), 591-602.
- Neto, F., & Furnham, A. (2011). Sex differences in parents' estimations of their own and their children's multiple intelligences: a Portuguese replication. *The Spanish journal of psychology*, 14(1), 99-110.
- Neto, F., Furnham, A., & da Conceição Pinto, M. (2009). Estimating One's Own and One's Relatives' Multiple Intelligence: A cross-cultural study from east Timor and Portugal. *The Spanish journal of psychology*, 12(2), 518-527.
- Neto, F., Ruiz, F., & Furnham, A. (2008). Sex differences in self-estimation of multiple intelligences among Portuguese adolescents. *High Ability Studies*, 19(2), 189-204.
- Ortner, T. M., Müller, S. M., & Garcia-Retamero, R. (2011). Estimations of parental and self intelligence as a function of parents' status: A cross-cultural study in Germany and Spain. *Social Science Research*, 40(4), 1067-1077.
- Pérez, L. F., González, C., & Beltrán, J. A. (2010). Parental estimates of their own and their relatives' intelligence. A Spanish replication. *Learning and individual differences*, 20(6), 669-676.
- Petrides, K. V., Furnham, A., & Martin, G. N. (2004). Estimates of emotional and psychometric intelligence: Evidence for gender-based stereotypes. *The Journal of social psychology*, 144(2), 149-162.
- Proyer, R. T. (2011). Being playful and smart? The relations of adult playfulness with psychometric and self-estimated intelligence and academic performance. *Learning and Individual Differences*, 21(4), 463-467.
- Rammstedt, B., & Rammsayer, T. H. (2000). Sex differences in self-estimates of different aspects of intelligence. *Personality and Individual Differences*, 29(5), 869-880.
- Rammstedt, B., & Rammsayer, T. H. (2002a). Gender differences in self-estimated intelligence and their relation to gender-role orientation. *European Journal of Personality*, 16(5), 369-382.
- Rammstedt, B., & Rammsayer, T. H. (2002b). Self-estimated intelligence: Gender differences, relationship to psychometric intelligence and moderating effects of level of education. *European Psychologist*, 7(4), 275.

- Reilly, J., & Mulhern, G. (1995). Gender differences in self-estimated IQ: The need for care in interpreting group data. *Personality and Individual Differences*, 18(2), 189-192.
- Sachs, J. (2004). Correlates of academic ability among part-time graduate students of education in Hong Kong. *Psychologia*, 47(1), 44-56.
- Siegle, D., Rubenstein, L. D. V., Pollard, E., & Romey, E. (2010). Exploring the relationship of college freshmen honors students' effort and ability attribution, interest, and implicit theory of intelligence with perceived ability. *Gifted Child Quarterly*, 54(2), 92-101.
- Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivations. *Sex Roles*, 50, 241-252.
- Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation?. *European journal of personality*, 22(3), 185-209.
- Steinmayr, R., & Spinath, B. (2009). What explains boys' stronger confidence in their intelligence?. *Sex Roles*, 61(9-10), 736-749.
- Stieger, S., Kastner, C. K., Voracek, M., Von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2010). Independent effects of personality and sex on self-estimated intelligence: Evidence from Austria. *Psychological reports*, 107(2), 553-563.
- Swami, V., Furnham, A., & Kannan, K. (2006). Estimating self, parental, and partner multiple intelligences: A replication in Malaysia. *The Journal of social psychology*, 146(6), 645-655.
- Syzmanowicz, A., & Furnham, A. (2013). Gender and gender role differences in self- and other-estimates of multiple intelligences. *The Journal of Social Psychology*, 153, 399-423.
- Van Soom, C., & Donche, V. (2014). Profiling first-year students in STEM programs based on autonomous motivation and academic self-concept and relationship with academic achievement. *PloS one*, 9(11), e112489.
- Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What makes you think you're so smart? Measured abilities, personality, and sex differences in relation to self-estimates of multiple intelligences. *Journal of Individual Differences*, 29(1), 35-44.
- von Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology*, 105, 243-253.
- Watson, J. M. (1989). An exploration of gender differences in tertiary mathematics. *Mathematics Education Research Journal*, 1(1), 1-10.

- Wouters, S., Germeijs, V., Colpin, H., & Verschueren, K. (2011). Academic self-concept in high school: Predictors and effects on adjustment in higher education. *Scandinavian journal of psychology*, 52(6), 586-594.
- Yuen, M., & Furnham, A. (2006). Sex differences in self-estimation of multiple intelligences among Hong Kong Chinese adolescents. *High Ability Studies*, 16(02), 187-199.

REFERENCES

- Ackerman, P. L. (2006). Cognitive sex differences and mathematics and science achievement. *American Psychologist*, 61(7), 722-723.
- Ackerman, P. L., & Wolman, S. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, 13, 57-78.
- Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, 49, 1621-1630.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Beloff, H. (1992). Mother, father, and me: Our IQ. *The Psychology*, 5, 309-311.
- Bennett, M. (1996). Men's and women's self-estimates of intelligence, *The Journal of Psychology*, 136(3), 411-412.
- Bennett, M. (2000). Self-estimates and population estimates of ability in men and women. *Australian Journal of Psychology*, 52, 23-28.
- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 42, 155-162.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88(4), 354-364.
- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59, 960-970.
- Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, 34(3), 139-153.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1-40.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Oxford, UK: John Wiley & Sons.
- Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues*, 28, 59-78.

- Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin*, 38, 209-219.
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, 106, 676–713.
- Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research*, 54(3), 427-456.
- Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, DC: American Psychological Association.
- Byrne, B. M., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125, 367-383.
- Chamorro-Premuzic, T., & Furnham, A. (2005). *Personality and Intellectual Competence*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chen, G., Gully, S. M., & Eden, D. (2004). General self-efficacy and self-esteem: Toward theoretical and empirical distinction between correlated self-evaluations. *Journal of Organizational Behavior*, 25, 375-395.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Colom, R., Juan-Espinosa, M., Abad, F., & Garcia, L. (2000). Negligible sex differences in general intelligence. *Intelligence*, 28, 57–68.
- Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, 57, 1082-1090.
- Eagly, A.H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, 54, 408-423.
- Eagly, A. H., & Wood, W. (2013). The nature-nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science*, 8, 340-357.
- Epstein, S. (1973). The self-concept revisited. *American Psychologist*, 28, 404-416.
- Erdwins, C., Small, A., & Gross, R. (1980). The relationship of sex role to self-concept. *Journal of Clinical Psychology*, 36(1), 111–115.
- Felson, R. B. (1981). Ambiguity and bias in the self-concept. *Social Psychology Quarterly*, 44, 64-69.

- Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138, 296-321.
- Furnham, A. (2001). Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31, 1381-1405.
- Furnham, A., & Akande, A. (2004). African parents' estimates of their own and their children's multiple intelligences. *Current Psychology*, 22, 281-294.
- Furnham, A., & Baguma, P. (1999). Self-Estimates of Intelligence: A Cross-Cultural Study from Three Continents. *North American Journal of Psychology*, 1(1), 69-78.
- Furnham, A., & Buchanan, T. (2005). Personality, gender, and self-perceived intelligence. *Personality and Individual Differences*, 39, 543-555.
- Furnham, A., & Bunclark, K. (2006). Sex differences in parents' estimations of their own and their children's intelligence. *Intelligence*, 34, 1-14. doi:10.1016/j.intell.2005.05.005
- Furnham, A., & Chamorro-Premuzic, T. (2005). Estimating one's own and one's relatives' multiple intelligence: A study from Argentina. *The Spanish Journal of Psychology*, 8(1), 12-20.
- Furnham, A., & Fukumoto, S. (2008). Japanese parents' estimates of their own and their children's multiple intelligences: Cultural modesty and moderate differentiation. *Japanese Psychological Research*, 50, 63-76.
- Furnham, A., Hosoe, T., & Li-Ping Tang, T. (2002). Male hubris and female humility? A cross-cultural study of ratings of self, parental, and sibling multiple intelligence in America, Britain, and Japan. *Intelligence*, 30, 101-115.
- Furnham, A., Moutafi, J., & Chamorro-Premuzic, T. (2005). Personality and intelligence: Gender, the big five, self-estimated and psychometric intelligence. *International Journal of Selection and Assessment*, 13(1), 11-24.
- Furnham, A., Rakow, T., Sarmany-Schuller, I., & De Fruyt, F. (1999). European differences in self-perceived multiple intelligences. *European Psychologist*, 4, 131-140.
- Furnham A., & Rawles, R. (1995). Sex differences in the estimation of intelligence. *Journal of Social Behavior and Personality*, 10(3), 741-748.
- Furnham, A., Shahidi, S., & Baluch, B. (2002). Sex and culture differences in perceptions of estimated multiple intelligence for self and family: A British-Iranian comparison. *Journal of Cross-Cultural Psychology*, 33, 270-285.

- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY: Basic Books.
- Gentile, B., Grabe, S., Dolan-Pascoe, B., Twenge, J. M., Wells, B. E., & Maitino, A. (2009). Gender differences in domain-specific self-esteem: A meta-analysis. *Review of General Psychology*, 13(1), 34.
- Guimond, S., Chatard, A., Martinot, D., Crisp, R., & Redersdorff, S. (2006). Social comparison, self-stereotyping, and gender differences in self-construals. *Journal of Personality and Social Psychology*, 90(2), 221-242. doi:10.1037/0022-3514.90.2.221
- Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1), 123-142.
- Hattie, J. (1992). *Self-concept*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research in gender differences. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 14–50). Baltimore, MD: Johns Hopkins University Press.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Hogan, H. W. (1978). Academic performance as a function of self-estimated intelligence and anxiety. *Indian Journal of Psychology*.
- Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 80(2), 327-334.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.
- Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 373-398.
- Hyde, J. S. (2016). Sex and cognition: gender and cognitive functions. *Current Opinion in Neurobiology*, 38, 53-56.

- Hyde, J. S., Fennema, E., & Lamon S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139-155
- Hyde, J. S., & Linn, M.C. (1988). Gender differences in verbal ability: a meta-analysis. *Psychological Bulletin*, 104, 53–69
- James, W. (1890/1981). *The principles of psychology*. Cambridge, MA: Harvard University Press.
- Kaufman, J. C. (2012). Self estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences*, 22, 118–222.
- Kessels, U. (2005). Fitting into the stereotype: how gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, 20, 309–323.
- Kornilova, T. V., Kornilov, S. A., & Chumakova, M. A. (2009). Subjective evaluations of intelligence and academic self-concept predict academic achievement: Evidence from a selective student population. *Learning and Individual Differences*, 19, 596-608.
- Kurtz-Costes, B., Copping, K. E., Rowley, S. J., & Kinlaw, C. R. (2014). Gender and age differences in awareness and endorsement of gender stereotypes about academic abilities. *European Journal of Psychology of Education*, 29(4), 603-618.
- Lindberg, S. M., Hyde, J. S., Petersen, J., & Linn M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, 136, 1123-1135.
- Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 147–158). New York, NY: Russell Sage Foundation.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage
- Maccoby, E. E., & Jacklin, C. N. (1978). *The psychology of sex differences* (Vol. 1). Stanford, CA: Stanford University Press.
- Marčič, R., & Kobal Grum, D. (2011). Gender differences in self-concept and self-esteem components. *Studia Psychologica*, 53, 373-384.
- Marsh, H.W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H.W. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, 24, 17-19.

- Marsh, H. W. (1990a). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2, 77-172.
- Marsh, H. W. (1990b). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82(4), 623-636.
- Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent self-concept. An interim test manual and research monograph*. Macarthur, New South Wales, Australia: University of Western Sydney, Faculty of Education
- Marsh, H.W. (1994). Using the National Longitudinal Study of 1988 to evaluate theoretical models of self-concept: The Self-Description Questionnaire. *Journal of Educational Psychology*, 86, 439-456.
- Marsh, H. W. (2006). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, UK: British Psychological Society.
- Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology* 80(3), 366-380.
- Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Eds.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 38-90). Oxford, UK: John Wiley & Sons.
- Marsh, H. W., & Hocevar, D. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. *Psychological Bulletin*, 97, 562-582.
- Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement*, 21(2), 153-174.
- Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychological*, 20(3), 107-123.
- Marsh, H. W., & Yeung, A. S. (1998). Top-down, bottom-up, and horizontal models: The direction of causality in multidimensional, hierarchical self-concept models. *Journal of Personality and Social Psychology*, 75(2), 509-527.
- Nasser, R., & Singhal, S. (2006). How youth in India and Lebanon rate their intelligence. *Journal of Social Sciences*, 2, 93-99.
- Nosek, B. A., Frederick, L. S., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... Greenwald, A. G. (2009). National differences in gender-science stereotypes

- predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26), 10593-10597.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193-203
- Peterson, E. R., & Whiteman, M. C. (2007). "I think I can, I think I can... ": The interrelationships among self-assessed intelligence, self-concept, self-efficacy and the personality trait intellect in university students in Scotland and New Zealand. *Personality and Individual Differences*, 43, 959-968.
- Pigott, T. D. (2012). *Advances in meta-analysis*. New York, NY: Springer.
- Rammstedt, B., & Rammsayer, T. H. (2002a). Gender differences in self-estimated intelligence and their relation to gender-role orientation. *European Journal of Personality*, 16(5), 369–382.
- Reilly, J., & Mulhern, G. (1995). Gender differences in self-estimated IQ: The need for care in interpreting group data. *Personality and Individual Differences*, 18(2), 189-192.
- Reilly, D., & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles*, 68, 521-535.
- Reynolds, W. M., Ramirez, M. P., Magriña, A., & Allen, J. E. (1980). Initial development and validation of the academic self-concept scale. *Educational and Psychological Measurement*, 40(4), 1013-1016.
- Roberts, T. A. (1991). Gender and the influence of evaluations on self-assessment in achievement settings. *Psychological Bulletin*, 109, 297-308.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Shavelson, R. J., & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology*, 74(1), 3-17.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3), 407-441.
- Shipstone, K., & Burt, S. (1973). Twenty-five years on: A replication of Flugel's (1947) work on lay popular views of intelligence and related topics. *British Journal of Educational Psychology*, 56, 183-187.

- Skaalvik, E. M. (1997). Issues in research on self-concept. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, 51-97). New York: JAI Press.
- Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivations. *Sex Roles, 50*, 241-252.
- Stake, J. E. (1992). Gender differences and similarities in self-concept within everyday contexts. *Psychology of Women Quarterly, 16*, 349-363.
- Sternberg, R. (1990). *Metaphors of mind: conceptions of the nature of intelligence*. Cambridge, UK: Cambridge University Press.
- Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88*(3) 397-407.
- Swann, W. B. Jr., Chang, Schneider, C. & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist, 62*(2), 84-94.
- Syzmanowicz, A., & Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and Individual Differences, 21*, 492-504
- Syzmanowicz, A., & Furnham, A. (2013). Gender and gender role differences in self- and other-estimates of multiple intelligences. *The Journal of Social Psychology, 153*, 399-423.
- Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90*, 334-349.
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.
- von Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology, 105*, 243-253.
- von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2009). Decomposing self-estimates of intelligence: Structure and sex differences across 12 nations. *British Journal of Psychology, 100*, 429-442.
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174.

- Wallentin, M. (2009). Punative sex differences in verbal abilities and language cortex: A critical review. *Brain & Language*, 108, 175-183.
- Wilgenbusch, T., & Merrell, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly*, 14(2), 101-120.
- Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In J. Olson & M. Zanna (Eds.), *Advances in experimental social psychology*, Vol. 46, (pp. 55-123). Burlington, MA: Academic Press.
- Workman, L. (2004). Self-perception of intelligence in male and female undergraduates in old and new Welsh universities. *Psychology Learning and Teaching*, 4, 22–26.
- Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist*, 70, 10-20.